REMARKS

Claims 1, 2, and 22-45 are currently pending. Claims 3-21 are canceled. Claims 23-29, 31, 33-39, 41 and 44-45 have been amended herein. Specific support for amendments to claims 24-29 and 34-39 can be found, for example, in Figures 21, 48 and 49. All the amended claims are fully supported by the instant specification and no new matter has been introduced.

Objections

The Examiner has objected to claims 24-29 and 34-39 under 37 CFR 1.821(d), asserting that a SEQ ID NO is required. These claims have been amended to reference the corresponding SEQ ID NO.

The disclosure is objected to by the Examiner for allegedly failing to comply with the requirements of 37 CFR 1.821(d), asserting that SEQ ID NOs are required for Figures 23-40. Replacement Figures 23-40 are provided herein with SEQ ID NOs inserted.

Rejections under 35 U.S.C. §112, first paragraph

Claims 1, 2, 22, 24-32, 34-43 and 45 are rejected by the Examiner under 35 U.S.C. §112, first paragraph, because the specification allegedly does not enable any human monoclonal antibody or binding portion thereof that binds to any Platelet Derived Growth Factor for any purpose. The Examiner states that the specification discloses only a human monoclonal antibody or antigen binding portion thereof that binds to human Platelet Derived Growth Factor D comprising a heavy chain amino acid sequence comprising SEQ ID NO:48 and a light chain amino acid sequence comprising SEQ ID NO:49. The Examiner states the factors to be considered in determining whether undue experimentation is required to practice the claimed invention are summarized In re Wands (858 F2d 731, 737, 8 USPQ2d 1400, 1404 (Fed Cir. 1988)) and include 1) the scope of the claim, 2) the amount of direction or guidance provided, 3) the lack of sufficient working examples, 4) the unpredictability in the art and 5) the amount of experimentation required to enable one of skill in the art to practice the claimed invention. The Examiner contends that the specification disclosure is insufficient to enable one skilled in the art to practice the invention as broadly claimed without an undue amount of experimentation.

Applicants strongly disagree with the Examiner's position.

Scope of the claims is enabled

The scope of the currently pending claims is a human monoclonal antibody that binds to Platelet Derived Growth Factor D (PDGFD), further encoded by or derived from specific VH (V_H1-8) and JH (V_H6B) genes. The claims **do not**, as the Examiner contends, pertain to human antibodies binding any PDGF, but pertain specifically to PDGFD. Contrary to the Examiner's contention that there is insufficient guidance about "other PDFGD", the present specification quite clearly defines PDFGD. The Examiner's attention is invited to the specification at page 14 line 25 to page 15 line 2:

> A novel PDGF, PDGF-D, has recently been cloned and characterized. See LaRochelle et al. Nature Cell Biology 3:517 (2001), GenBank Accession No. AF335584, International Patent Application No. WO 01/25433, USSN 60/158,083, filed October 7, 1999; USSN 60/159,231, filed October 13, 1999; USSN 60/174,485 filed January 4, 2000; USSN 60/186,707 filed March 3, 2000; USSN 60/188,250, filed March 10, 2000; USSN 60/223,879, filed August 8, 2000; USSN 60/234,082, filed on September 20, 2000; USSN 09/685,330, filed on October 5, 2000; PCT Application US00/27671, filed October 6, 2000; USSN 09/688,312, filed October 13, 2000 and USSN 09/715,332, filed November 16, 2000.

The Examiner's attention is further directed to Figures 1 and 2 as referenced in the specification at page 15, lines 11-12 which provide the nucleic acid and amino acid sequence of PDGFD. PDGFD is specifically disclosed and described. The specificity of the antibodies of the invention to PDGFD, not *any PDGF* is clear, defined and enabled.

The Examiner relies upon Ngo et al, seemingly to support the Examiner's argument that the immunogen, and therefore the specificity of the antibodies of the invention are not defined, therefore contending antibodies to PDGFD are not enabled. The Examiner states that:

> "Ngo et al, teach that the amino acid positions within the polypeptide/protein that can tolerate change such as conservative substitution or no substitution, addition or deletion which are critical to maintain the protein's structure/function will require guidance.

However, Applicants respectfully point out that Ngo et al is concerned specifically with the ability, or lack thereof to *computationally* predict the tertiary structure of a protein based upon an amino acid sequence. For the current invention, computational prediction is not necessary nor is it relevant. Ngo et al states:

> "A protein molecule is a covalent chain of amino acid residues. Although it is topologically linear, in physiological conditions it folds into a unique (though flexible) three-demensional structure...referred to as the native structure."

Furthermore Ngo et al states:

> "It is not known whether there exists an efficient algorithm for predicting the structure of a given protein from its amino acid sequence alone. Decades of research have failed to produce such an algorithm, yet Nature seems to solve the problem. Proteins do fold!"

Given a nucleotide or amino acid sequence, it is well known in the art how to make a protein in a biological system. For example, the specification teaches expressing PDGFD in HEK293 cells to obtain protein for immunization (see Example 1, page 48). The immunogen is defined, and more pertanently the antibodies of the invention are enabled and Ngo offers nothing to refute this.

The Examiner further relies upon Kuby et al, who the Examiner contends teach:

> "antibody epitopes (B cell epitopes) are not linear and are comprised of complex three-dimensional array of scattered residues which will fold into specific conformation that contribute to binding. Immunization with a peptide fragment derived from a full-length polypeptide may result in antibody specificity that differs from the antibody specificity directed against the native full-length polypeptide."

The Examiner concludes that:

> "Without the specific amino acid residues, it is unpredictable which undisclosed immunogen would generate human monoclonal antibody that binds to all PDGFD."

While the Examiner may have characterized the teachings of Kuby et al, Applicants respectfully submit that these teachings are **not** relevant to the current invention. The invention does not pertain to B cell epitopes. Furthermore, the claims are not directed to methods of generating antibodies to PDGFD. Claims are directed to human antibodies that bind to PDGFD. Additionally, the antigen used to generate the human antibodies of the present invention is disclosed in the specification (see Example 1, page 48). Specifically the gene product of HEK293 cells transfected with PDGFD produces an active fragment "p35" which was used as the immunogen. PDGFD p35 is known in the art (see *See* LaRochelle et al. *Nature Cell Biology* 3:517 (2001) which is referenced in the specification at page 14, lines 25-26, copy attached as Attachment A). Breifly, the reference describes cloning of PDGFD cDNA into a mammalian expression vector and subsequent transfection into HEK293 cells. When PDGFD is purified

from serum-containing conditioned medium, a principal species, p35, is obtained. Sequence analysis of the amino terminus of p35 shows it is the product of proteolytic cleavage after R247 or R249. PDGFD is secreted as a dimer and is proteolytically processed in the presence of sera to p35. Hence Applicants are not relying on a "peptide fragment" in the way that Kuby refers to "fragments" but rather to a biologically produced end product as purified from conditioned culture media after protein synthesis and secretion. The antibodies of the invention are enabled and Kuby offers nothing to even suggest otherwise.

The Examiner further contends that the claims encompass any number or combination of heavy and light chains, heavy chain CDRs and light chain CDRs and that there is insufficient guidance as to which combinations in an antibody would maintain the same binding specificity. The Examiner states that the function of an antibody is dependent on its three dimensional structure and that amino acid sequence changes may adversely affect antibody activity. Even fragments of an antibody may not retain binding activity according to the Examiner. Additional concerns regarding what residues comprise CDRs, critical framework residues that affect antibody binding are raised by the Examiner as lacking guidance in the present specification. The Examiner also relies upon Janeway et al., to support the Examiner's position that there is insufficient guidance provided in the specification.

Applicants strongly disagree. The Examiner states that Janeway et al teach the association of different heavy and light chain variable regions from the binding site. Applicants respectfully submit that there is no teaching in Janeway (pages 3:1 to 3:4 provided by the Examiner) that remotely supports the Examiner's contention.

Applicants strongly disagree that there is insufficient guidance provided as preferred germline sequences as well as 19 specific combinations specifically binding PDGFD are provided in the specification (see for example Figures 48 and 49). Also CDRs and framework sequences are also described in the specification (see for example Figures 48-57). Furthermore, *it is well known in the art* how to create, for example libraries of various combinations of heavy and light chains and/or specific CDRs and then how to screen such libraries for those combinations that bind to an antigen. For example, over a decade ago, Marks et al described taking a sequence encoding a CDR and introducing it into a repertoire of variable domains lacking the respective CDR (e.g., CDR3), using recombinant DNA technology, displaying the repertoire in a suitable host system such as the phage display system and selecting for antigen-

binding members of the repertoire (*Bio/Technology (1992) 10: 779-783*, provided as Attachment B). Analogous shuffling or combinatorial techniques are also well known in the art and may be used (e.g. *Stemmer, Nature (1994) 370: 389-391,* provided as Attachment C).

The Examiner's rejection of the claims based upon some of the combinations not maintaining PDGFD binding specificity is incorrect as a matter of law. The Examiner acknowledges that the specification would enable the skilled artisan to practice some of the disclosed embodiments of the claimed invention, but that others would not be enabled. A claim may, however, encompass inoperative embodiments and still meet the enablement requirement of 35 USC 112, first paragraph (see Atlas Powder Co. v. E.I. Du Pont De Nemours & Co., 750 F. 2d 1569, 1576, 224 USPQ 409, 413 (Fed Cir. 1984), In re Angstadt, 537 F.2d 498, 504, 190 USPQ 213, 218 (CCPA 1976), In re Cook, 439 F.2d 730, 732, 169 USPQ 298, 300 (CCPA 1971).

Applicants also disagree with the Examiner's rejection of claims 24-29 and 34-39 for the use of the term "comprising". These claims each pertain to a specific CDR that is comprised by the human antibody. The antibody, in most cases will also have other CDR regions, framework regions as well as optionally constant domains. Therefore the open-ended language correctly reflects the invention, "the antibody comprises a heavy or light chain CDR region...".

Applicants respectfully submit that the scope of the claims is enabled by the present specification.

Direction and guidance are adequately provided

The Examiner contends that the specification provides insufficient guidance as to which particular VH1-8 family gene, JH6B family gene, and D5-18 family gene encode the antibody of claims 22, 30, 32 and 42. Applicants disagree, these terms and specifically these genes are well known in the art, see generally Kabat Sequences of Proteins of Immunological Interest, National Institutes of Health, Bethesda, Md. 1987 and 1991. Also, please see the specification at page 62 line 30 to page 63 line 3 where the protein sequence of VH1-8 is provided. Furthermore see Figure 50 which shows 7 exemplary VH1-8 derived sequences. The J region JH6B was described in Rabbitts, T. H. (1983) (*Biochem. Soc. Trans., 11, 119-126.*, provided as Attachment D). The D region D5-18 was described by Corbett et al (1997) (*J. Mol. Biol., 270, 587-597,* provided as Attachment E).

The MPEP states:

"not everything necessary to practice the invention need be disclosed. In fact, what is well-known is best omitted. In re Buchner, 929 F.2d 660, 661, 18 USPQ2d 1331, 1332 (Fed. Cir. 1991)." MPEP 2164.08

The Examiner states that the specification provides insufficient guidance as to the introns as well as exons encoding an antibody of the invention. Applicant disagrees as introns do not code for proteins, and the claims specifically pertain to a protein (antibody) encoded by a human VH1-8, JH6B and D5-18 family gene. Therefore details of genomic introns are irrelevant to enablement of the invention.

Applicants respectfully submit that adequate direction and guidance supporting the invention is provided. Furthermore an antibody encompassed by claims 22, 30, 32 and 42 is, in fact, enabled in the present specification.

Sufficient working examples are provided

The current invention provides the nucleotide and amino acid sequence of such affinity matured V domain FR and CDR regions that bind PDGFD. The application discloses no less than **19** human monoclonal antibodies that bind to PDGFD. These antibodies include **7** human monoclonal antibodies that bind to PDGFD encoded by or derived from human VH1-8; **13** human monoclonal antibodies that bind to PDGFD encoded by or derived from human JH6B; and **3** human monoclonal antibodies that bind to PDGFD encoded by or derived from human D5-18.

Given the present teaching, the unpredictability of the art is low and undue experimentation by one skilled in the art is not required to practice the claimed invention.

Given the teachings of the specification, Applicants respectfully submit that one of skill in the art would recognize that the present invention identifies the specific germline human antibody heavy chain V, D, J combinations and light chain V, J combinations that provide for binding to PDGFD. Furthermore, the current invention teaches 19 examples of affinity matured antibodies that bind PDGFD. The level of skill in the art of antibody engineering is quite high. Provided with the applicant's teachings, particularly specific V, D, J genes, affinity matured sequences, nucleic acids encoding them and amino acid sequences, one of skill in the art would

know how to make the antibodies of the invention as well as generate various combinations, for example various combinations of the identified V, D, and J regions or CDRs, utilizing the sequences of the invention using no more than routine experimentation and techniques well known in the art, including recombinant and synthetic methods (*Maniatis (1990) Molecular Cloning, A Laboratory Manual, 2.sup.nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., and Bodansky et al. (1995) The Practice of Peptide Synthesis, 2.sup.nd ed., Spring Verlag, Berlin, Germany*). Shuffling or combinatorial techniques are well known in the art and may be used (e.g. *Stemmer, Nature (1994) 370: 389-391*). Furthermore, as is understood by those of skill in the art, novel H variable or L variable regions may be generated derived from the sequences taught in the specification using random mutagenesis of one or more selected H variable and/or L variable genes, such as error-prone PCR (*Gram et al. Proc. Nat. Acad. Sci. U.S.A. (1992) 89: 3576-3580,* provided as Attachment F). Another method that may be used is to direct mutagenesis to CDRs of H variable or L variable genes (*Barbas et al. Proc. Nat. Acad. Sci. U.S.A. (1994) 91: 3809-3813* provided as Attachment G*; Schier et al. J. Mol. Biol. (1996) 263: 551-567,* provided as attachment H). Similarly, one or more, or all three CDRs may be grafted into a repertoire of H variable or L variable domains, which are then screened for an antigen-binding fragment specific for PDGFD.

Applicants submit that the present invention is clearly enabled, the scope of the claims in keeping with the scope of the discovery and disclosure, sufficient guidance and numerous working examples are provided that are commensurate with the level of skill in the art such that one of skill would be able to practice the invention as claimed. The rejection of claims 1-2, 22, 24-32, 34-43 and 45 under 35 U.S.C. §112 first paragraph should be withdrawn.

Rejections under 35 U.S.C. §112, first paragraph

The Examiner has rejected claims 1-2, 22, 24-32, 34-43 and 45 under 35 U.S.C. §112, first paragraph as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor, at the time the application was filed, had possession of the claimed invention. The Examiner contends that "the specification does not reasonably provide a written description of *any* human monoclonal

antibody that binds to *any* Platelet Derived Growth Factor such ...as set forth in claims 1-2, 22, 24-32, 34-30 (sic), 32-40" (emphasis added).

Applicants strongly disagree. The claims pertain to a human monoclonal antibody that binds to Platelet Derived Growth Factor D, **not** "any" Platelet Derived Growth Factor. The claims further pertain to certain human monoclonal antibodies comprising a specific heavy chain amino acid sequence (claim 1); a specific light chain amino acid sequence (claim 2); are encoded by or derived from specific germline sequences VH1-8 and JH6B (claims 22, 32); and have specific heavy and light chain amino acid sequences (claims 23, 33); have specific heavy chain CDR1 (claims 24, 34), CDR2 (claims 25, 35), CDR3 (claims 26, 36) or light chain CDR1 (claims 27, 37), CDR2 (claims 28, 38), CDR3 (claims 29, 39) or further comprise D region D5-18 (claims 30, 40). The specification absolutely provides specific written description of human monoclonal antibodies that bind to PDGFD. Nineteen specific examples are provided. Also provided is the analysis of the germline sequences utilized in the *in vivo* development of those antibodies which clearly defines the preferred germline sequences to use for obtaining antibodies that bind to PDGFD and such are clearly provided in the specification and pending claims. The specification absolutely provides specific written description of human monoclonal antibodies of the sequences defined in the pending claims. Therefore the rejection of claims 1-2, 22, 24-32, 34-43 and 45 under 35 U.S.C. §112, first paragraph should be withdrawn. Such action is respectfully requested.

Rejections under 35 U.S.C. §112, second paragraph

The Examiner has rejected claims 23-29, 31-39, 41, 44 and 45 under 35 U.S.C. §112, second paragraph as being indefinite. Specifically, the Examiner has rejected claims 23-29, 31-39 and 44 as indefinite for the use of the term "having". Claims have been amended to recite "comprising". Claims 31, 41 and 45 are rejected for lacking antecedent basis for "detectable marker". These claims have been amended to recite that said monoclonal antibody further comprises a detectable marker. Applicant respectfully requests that the rejection of Claims 23-29, 31-39, 41, 44 and 45 therefore be withdrawn.

Rejections under 35 U.S.C. §103(a)

Claims 22, 30-32, 40-43 and 45 are rejected by the Examiner under 35 U.S.C. §103(a) as being unpatentable over US Patent 6, 706, 687 in view of Green et al. The Examiner asserts that the '687 patent teaches monoclonal antibody and antigen binding portion thereof that binds to platelet derived growth factor D (PDGFD) and further teaches the antibody labeled with a detectable marker and use of the antibody as an inhibitor or agonist of PDGFD or for diagnostic assays. As stated by the Examiner the claimed invention differs from the '687 patent in that the claimed invention is a human monoclonal antibody and furthermore, that it is encoded by specific human V, J and D genes. The '687 does not provide any teaching of what genes or sequences will provide an antibody with specificity to PDGFD.

The Examiner states that Green et al teach

> "a method of making human antibody that binds to any antigen of interest wherein the reference antibody is encoded by or derived from a one (sic) of the human VH1-8 family agene (sic) such as VH6 and a JH6B family of (sic) gene such as J6 family gene. The reference antibody further comprises a sequence encoded by a human D family of gene such as TGGTTATTAC."

Applicants respectfully disagree with the Examiner's characterization of Green's teaching. It is generally known that antibodies are produced in nature; *in vivo* by B lymphocytes and that each clone of B cells produces an antibody with an antigen receptor having a unique prospective antigen binding structure. The repertoire of antigen receptors, approximately 10^7 possibilities, exists *in vivo* prior to antigen stimulation. This diversity is produced by somatic recombination-the joining of different antibody gene segments. The selection of specific V, C and J regions (and D for the heavy chain) from amongst the various gene segments available (45 heavy chain V; 35 kappa V; 23 heavy chain D; 6 heavy chain J; 5 kappa J) generates approximately 10^{11} possible specificities from germline sequences. Upon exposure to antigen, particular B cells with antigen binding specificity based on germline sequences are activated, proliferate, and differentiate to produce antibodies of different isotypes as well as undergo somatic mutation and/or affinity maturation to produce antibodies of higher affinity for the antigen.

Green et al teaches a mouse deficient in mouse antibody production, but capable of producing human antibodies based upon human heavy and kappa light chain genes that are

introduced into the mouse germline using the yeast artificial chromosome approach. Green et al are concerned with engineering a mouse system that can produce the diverse repertoire of human V, D, J combinations that in turn produce the huge diversity of human antibody specificities. Green does not teach which V to combine with which J and/or D gene to provide a certain specificity of antibody. Green does not teach which V, J or D gene will provide an antibody that binds to PDGFD.

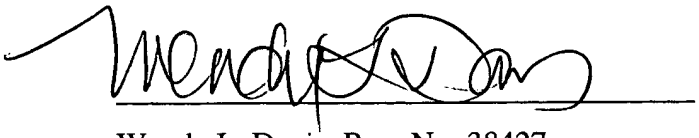
Applicants respectfully submit that the present invention has identified specific germline human antibody heavy chain V, D, J combinations and light chain V, J combinations including nucleotide and amino acid sequence of the VH and VL domain FR and CDR regions that bind PDGFD from a diverse repertoire. VH1-8 is a specific heavy chain V gene, described on page 62, line 30 to page 63, line 3 of the current specification. JH6B is a specific heavy chain J gene, known in the art and described by Rabbitts, T. H. (1983). Biochem. Soc. Trans., 11, 119-126. D5-18 is a specific heavy chain D region known in the art and described by Corbett et al (1997). J. Mol. Biol., 270, 587-597. Until the applicants' invention was made, it would not have been known to specifically use VH1-8, VH6b and D5-18 to provide for a human antibody that binds to PDGFD.

Applicants submit that the '687 merely suggests an antibody to PDGFD and Green merely provides a model for generating human antibodies. The combination of these two references in no way makes the Applicants invention obvious and the rejection of Claims 22, 30-32, 40-43 and 45 under 35 U.S.C. §103(a) should be withdrawn.

CONCLUSION

Applicant respectfully requests that the amendments and remarks made herein be entered and made of record in the file history of the present application. Applicant respectfully submits that this paper is fully responsive and that the pending claims are in condition for allowance. Such action is respectfully requested. If there are any questions regarding these amendments and remarks, the Examiner is encouraged to contact the undersigned at the telephone number provided below.

Respectfully submitted,



Date: November 19, 2004

Wendy L. Davis, Reg. No. 38427
Agent for Applicant
CuraGen Corporation
555 Long Wharf Drive, 9th Floor
New Haven, CT 06511
Direct: 203 974-6310
Main: 203 401-3330
Fax: 203 401-3351

Attachment A

PDGF-D, a new protease-activated growth factor

William J. LaRochelle*†, Michael Jeffers*, William F. McDonald*, Rajeev A. Chillakuru*, Neill A. Giese‡, Nathalie A. Lokker‡, Carol Sullivan‡, Ferenc L. Boldog*, Meijia Yang*, Corine Vernet*, Catherine E. Burgess*, Elma Fernandes*, Lisa L. Deegler*, Beth Rittman*, Juliette Shimkets*, Richard A. Shimkets*, Jonathan M. Rothberg* and Henri S. Lichenstein*

*CuraGen Corporation, 555 Long Wharf Drive, New Haven, Connecticut 06511, USA, and 322 East Main Street, Branford, Connecticut 06405, USA
‡COR Therapeutics, 256 East Grand Avenue, South San Francisco, California 94080, USA
†e-mail: wlarochelle@curagen.com

Platelet-derived growth factor (PDGF) has been directly implicated in developmental and physiological processes[1–5], as well as in human cancer, fibrotic diseases and arteriosclerosis[6]. The PDGF family currently consists of at least three gene products, PDGF-A, PDGF-B and PDGF-C, which selectively signal through two PDGF receptors (PDGFRs) to regulate diverse cellular functions. After two decades of searching, PDGF-A and B were the only ligands identified for PDGFRs. Recently, however, database mining has resulted in the discovery of a third member of the PDGF family, PDGF-C[7,8], a functional analogue of PDGF-A that requires proteolytic activation. PDGF-A and PDGF-C selectively activate PDGFR-α[7,9,10], whereas PDGF-B activates both PDGFR-α and PDGFR-β[9,11]. Here we identify and characterize a new member of the PDGF family, PDGF D, which also requires proteolytic activation. Recombinant, purified PDGF-D induces DNA synthesis and growth in cells expressing PDGFRs. In cells expressing individual PDGFRs, PDGF-D binds to and activates PDGFR-β but not PDGFR-α. However, in cells expressing both PDGFRs, PDGF-D activates both receptors. This indicates that PDGFR-α activation may result from PDGFR-α/β heterodimerization.

We searched a database of human expressed genes and identified a unique complementary DNA, the open reading frame (ORF) of which was predicted to encode a new member of the PDGF/vascular endothelial growth factor (VEGF) family, hereafter referred to as PDGF-D. Human PDGF-D (Fig. 1a) encodes a protein of 370 amino acids (relative molecular mass (M_r) 42,848, pI 7.88). Isolation of the murine PDGF-D cDNA and analysis of its predicted ORF (370 amino acids, M_r 42,808, pI 7.53) showed that PDGF-D is conserved across species (85% overall amino-acid identity; data not shown). Computer analysis using PSORT[12], PFAM[13] and PROSITE[14] predicted that human PDGF-D contains a characteristic signal peptide (amino acids 1–23), a CUB (InterPro 000859) domain[15] (amino acids 53–167), a PDGF/VEGF-homology domain (amino acids 272–362) and an *N*-linked glycosylation site (amino acid 276). BLASTP analysis revealed that human PDGF-D is most closely related to human PDGF-C, PDGF-B and PDGF-A (42%, 27% and 25% overall amino-acid identity, respectively; data not shown). An alignment of the core PDGF domains from PDGF-D (human and mouse), PDGF-C, PDGF-B and PDGF-A is presented in Fig. 1b. From this, it is apparent that PDGF-D retains seven of the eight invariant cysteine residues that are involved in intrachain and interchain disulphide bonding[16], with a substitution of a glycine residue at the site of the fifth invariant cysteine.

Using genomic DNA sequences obtained from GenBank, we also elucidated the exon/intron organization of PDGF-D. Analysis of the genomic DNA sequence indicated that the PDGF-D gene is comprised of seven exons (Fig. 1a), like PDGF-A and PDGF-B. In PDGF-D, both the CUB (exons two and three) and PDGF (exons six and seven) domains span two exons. PDGF-D lacks the carboxy-terminal retention motif found in the PDGF-A exon-six splice variant and in PDGF-B[17]. We identified an in-frame stop codon 9 base pairs (bp) upstream of the initiator methionine. We found that the chromosomal location of the PDGF-D gene is 11q22.3; we further pinpointed it by radiation hybrid analysis, which showed that the gene is located 3.36 cR from marker WI-9345 on chromosome 11 (data not shown).

We examined the expression profile of PDGF-D messenger RNA using a real-time quantitative polymerase chain reaction (PCR)[18]. In the 37 normal human tissues examined, PDGF-D was most highly expressed in the adrenal gland (Fig. 2a). Moderate levels of PDGF-D were present in pancreas, adipose tissue, heart, stomach, bladder, trachea, mammary gland, ovary and testis. In contrast, PDGF-B was highly expressed in heart, brain (substantia nigra), fetal kidney and placenta. Moderate expression levels were observed in brain (hippocampus), skeletal muscle, kidney and lung (Fig. 2a). PDGF-D transcripts were also highly expressed in some tumour cell lines (derived from glioblastomas, carcinomas and melanomas) and in some human cancer tissues (kidney and ovarian carcinoma; data not shown). These results show that PDGF-D has a localization that is distinct from that of PDGF-B.

To examine the biochemical properties of PDGF-D, we subcloned the PDGF-D cDNA into a mammalian expression vector. After transfection into 293 HEK cells and growth in the presence or absence of 10% FBS, western blotting (after SDS–polyacrylamide-gel eletrophoresis (SDS–PAGE) under reducing conditions) showed that PDGF-D was secreted as a species of M_r 49,000 (49K) or 20K, respectively (data not shown). We therefore expressed PDGF-D in the presence or absence of FBS and purified it to >95% homogeneity for biological characterization. As shown in Fig. 2b (lane 2), expression of PDGF-D under serum-free conditions resulted in the detection of the expected M_r 49K gene product under reducing conditions. A dimeric species, p84, was detected under non-reducing conditions (Fig. 2b, lane 1). When PDGF-D was purified from serum-containing conditioned medium, a principal species, p35, was observed, as well as a minor species, p15 (Fig. 2b, lane 3). Sequence analysis of the amino terminus of p35 revealed proteolytic cleavage after R247 or R249. Under reducing conditions, this product was found to migrate as three predominant species (p20, p14 and p6; Fig. 2b, lane 4). Further sequence analysis showed that p14 and p6 resulted from proteolytic cleavage of p20 at K339 or R340. As the majority of p35 migrated as a single band under non-reducing conditions, it is likely that the p35 dimer is nicked and that the chains are held together by disulphide bonds. These results indicate that PDGF-D is secreted as a dimer, which we have termed PDGF-DD, and that it is proteolytically processed in the presence of FBS.

© 2001 Macmillan Magazines Ltd

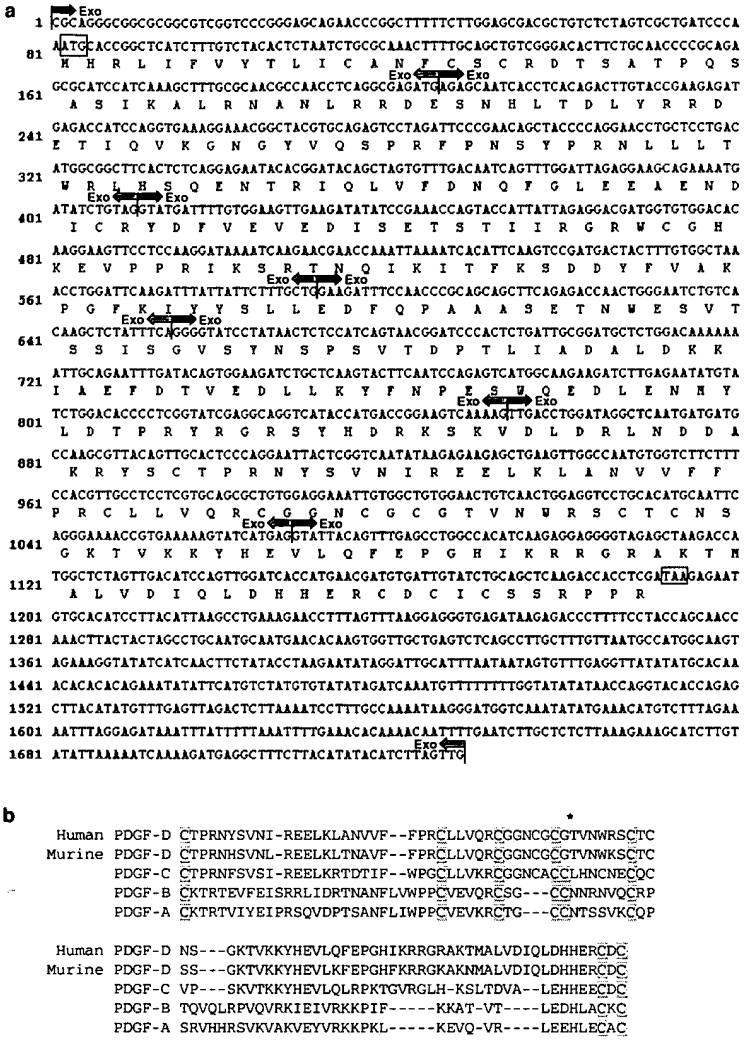


Figure 1 **Nucleotide and deduced amino-acid sequence of human PDGF-D.** **a**, The genomic structure and deduced encoded sequence of the human PDGF-D gene. Initiation and stop codons are boxed; intron/exon boundaries are indicated by arrows. Numbers represent nucleotide numbers. **b**, Comparison of core PDGF domains among PDGF family members. GenBank accession numbers used for alignment are AF335584 (human PDGF-D), AF335583 (murine PDGF-D), AAF80597 (PDGF-C), P01127 (PDGF-B) and P04085 (PDGF-A). Invariant cysteine residues are shaded; asterisk indicates the missing invariant cysteine residue in PDGF-D.

We tested recombinant PDGF-DD p84 and p35 for their ability to induce DNA synthesis in a 5-bromodeoxyuridine (BrdU)-incorporation assay. Using NIH 3T3 embryonic lung fibroblasts, p35 induced BrdU incorporation at a half-maximal concentration of 20 ng ml^{-1} (Fig. 2c). p84 did not induce BrdU incorporation at concentrations up to 1 μg ml^{-1} (Fig. 2c), nor did it block p35- or PDGF-BB-induced BrdU incorporation (data not shown). Comparatively, PDGF-AA and PDGF-BB induced half-maximal DNA synthesis at ~4 and 5 ng ml^{-1}, respectively. At concentrations of PDGF-DD of >10 ng ml^{-1}, there was an increase in overall BrdU incorporation compared with PDGF-AA and PDGF-BB. However, in other experiments that measured mitogenicity, PDGF-DD supported the growth of NIH 3T3 cells, CCD 1070 foreskin fibroblasts, MG-63 osteosarcoma cells, U118-MG glioblastoma cells and primary smooth-muscle cells to a similar extent to PDGF-BB (data not shown). Together, these results indicate that PDGF-DD may be a latent growth factor, the activity of which may be dependent on proteolytic dissociation of the PDGF/VEGF core domain from the CUB-containing region.

To investigate the possibility that PDGF-DD exerts its biological effects through PDGFR-α and/or PDGFR-β, we examined PDGFR binding and phosphorylation in well-characterized PDGFR-null cells engineered to express PDGFR-α or PDGFR-β[9,19]. As shown in Fig. 3a, PDGF-DD did not compete with ^{125}I-labelled PDGF-AA for binding to PDGFR-α in cells overexpressing this receptor (32DαR), at concentrations up to 250 nM. However, PDGF-DD did compete with ^{125}I-labelled PDGF-BB for binding to PDGFR-β in cells overexpressing this receptor (HR5βR), although higher concentrations of PDGF-DD were required relative to the PDGF-BB competitor (Fig. 3b). As expected, PDGF-AA did not compete with ^{125}I-labelled PDGF-BB for binding to PDGFR-β, confirming the specificity of the binding assay.

We further confirmed these results by analysing PDGFR activation using a two-site, enzyme-linked immunosorbent assay (ELISA) that quantitatively measures incorporation of phosphotyrosine into PDGFR-α or PDGFR-β. As shown in Fig. 3c, exposure of 32DαR cells for 10 min to PDGF-AA or PDGF-BB induced a

© 2001 Macmillan Magazines Ltd

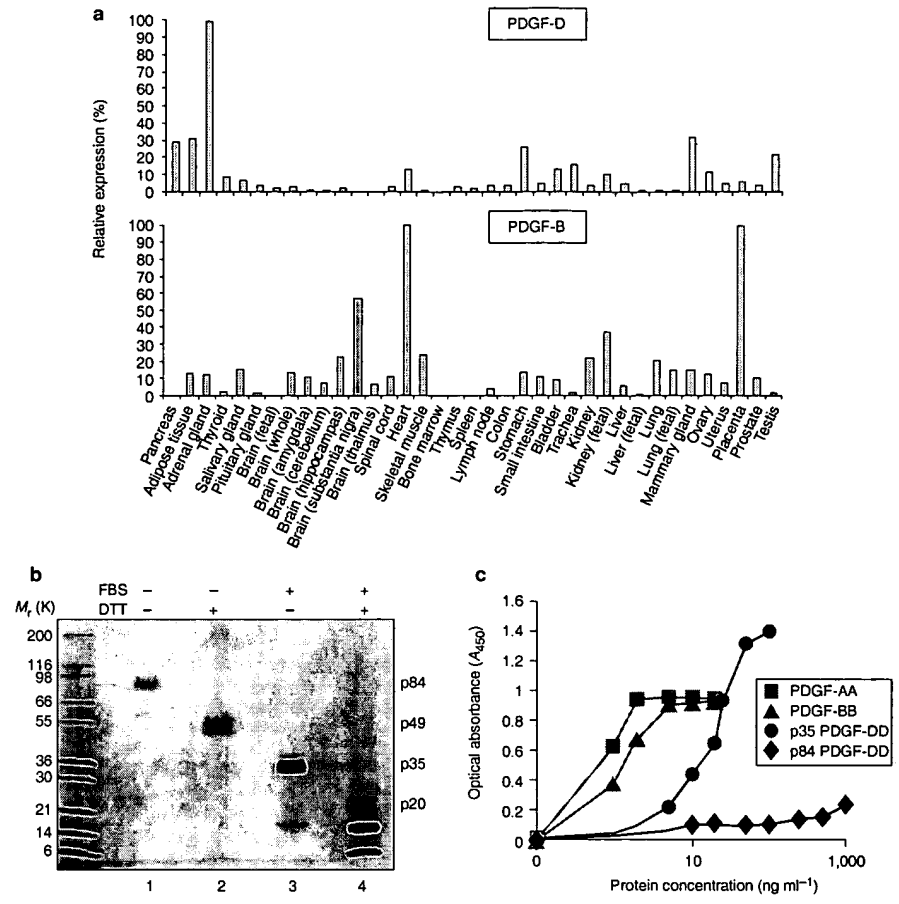


Figure 2 ***TaqMan* expression analysis, purification and biological activity of PDGF-DD. a**, Real-time quantitative PCR analysis of PDGF-DD expression. Normal tissues were analysed using specific *TaqMan* reagents for PDGF-D (upper panel) or PDGF-B (lower panel) derived from the indicated samples. Equal quantities of normalized RNA were used as a template in PCR reactions to obtain threshold cycle (C_T) values. C_T is expressed as per cent expression relative to the sample exhibiting the highest level of expression. **b**, Purification of PDGF-D. PDGF-D was purified from transfected HEK 293 cells cultured in the presence (lanes 3, 4) or absence (lanes 1, 2) of FBS. PDGF-D was resolved by SDS–PAGE and stained with Coomassie blue. Samples were treated with (+) or without (–) dithiothreitol (DTT). The positions of relative molecular mass markers are indicated on the left. **c**, BrdU-incorporation assay. NIH 3T3 fibroblasts were serum-starved, incubated with PDGF-DD p35 (circles), PDGF-DD p84 (diamonds), PDGF-BB (triangles) or PDGF-AA (squares) for 18 h, and assayed for incorporation of BrdU (s.d. did not exceed 15%).

four to ten-fold induction in tyrosine phosporylation of PDGFR-α. No induction was observed with PDGF-DD. In HR5βR cells, PDGF-BB and PDGF-DD, but not PDGF-AA, induced phosphotyrosine incorporation (Fig. 3d). PDGF-DD-induced phosphorylation was detected at concentrations as low as 10 ng ml^{-1}, but never reached the level of PDGF-BB-induced phosphorylation at the highest concentrations tested. Together, these results demonstrate that in cells expressing individual PDGFRs, PDGF-DD binds to and activates PDGFR-β, but not PDGFR-α.

We next measured PDGFR activation in CCD1070 fibroblasts expressing both PDGFR-α and PDGFR-β. As expected, PDGF-AA induced tyrosine phosphorylation of PDGFR-α, whereas PDGF-BB activated both PDGFR-α and PDGFR-β. Unexpectedly, PDGF-DD induced phosphotyrosine incorporation in both PDGFR-α and PDGFR-β (Fig. 3e, f). Similar results were obtained using MG-63 cells containing both PDGFRs (data not shown). We confirmed the CCD1070 results by PDGFR immunoprecipitation followed by anti-phosphotyrosine western blotting (Fig. 3g). Consistent with the results of the two-site ELISA, PDGF-DD and PDGF-BB activated both PDGFR-α and PDGFR-β. PDGF-AA induced phosphorylation only of PDGFR-α, which again confirms the specificity of the assay. p84 PDGF DD induced no PDGFR activation (data not shown).

To explain PDGFR-α activation, we investigated whether PDGF-DD binds directly to PDGFR-α in CCD1070 cells after binding to PDGFR-β. This could occur, for example, if interaction with PDGFR-β induces conformational changes in PDGF-DD that facilitate interaction with PDGFR-α. In the CCD1070 binding assay (Fig. 3h), PDGF-BB and PDGF-AA, but not PDGF-DD, blocked binding of ^{125}I-labelled PDGF-AA. PDGF-DD had no effect on binding of ^{125}I-labelled PDGF-AA at concentrations of >40 nM. We further confirmed the dependence of PDGF-DD signalling on interaction with PDGFR-β by blocking it with an antibody that neutralizes PDGFR-β, but not PDGFR-α (data not shown). The results were consistent with the 32DαR binding assay and indicate that PDGF-DD does not undergo PDGFR-β-induced conformational changes that promote competitive PDGFR-α binding.

We next investigated whether PDGFR-α is activated through interaction with PDGFR-β in PDGFR-α/β heterodimers. As shown in Fig. 3i, at concentrations selected for maximal phosphotyrosine incorporation, PDGF-BB and PDGF-DD (to a fivefold lesser extent) were able to induce formation of PDGFR-α/β heterodimers, whereas

© 2001 Macmillan Magazines Ltd

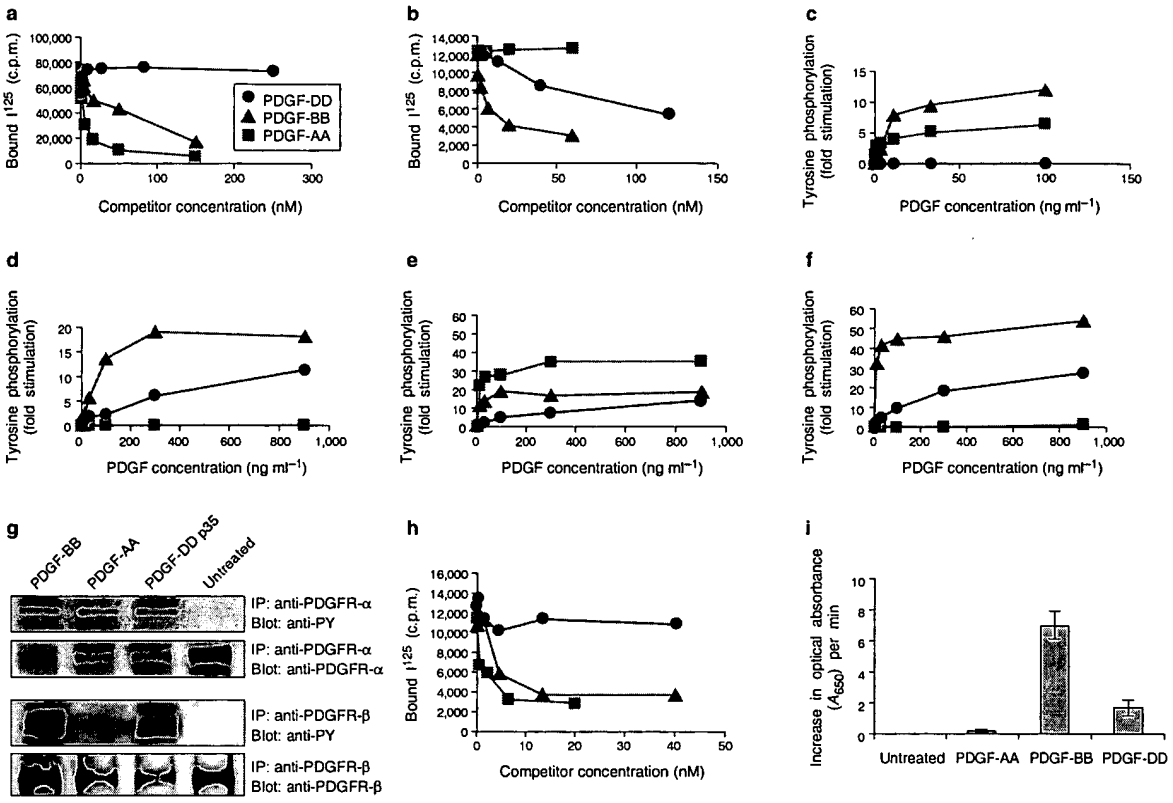


Figure 3 **PDGF-DD-binding, tyrosine-phosphorylation and heterodimerization of PDGFRs. a, b,** PDGFR binding. Competition by PDGF-DD (circles) for binding of ^{125}I-labelled PDGF-AA to 32DαR cells expressing PDGFR-α (**a**) or of ^{125}I-labelled PDGF-BB to HR5βR cells expressing PDGFR-β (**b**). Competition by PDGF-AA (squares) and PDGF-BB (triangles) are shown for comparison (s.d. did not exceed 15%). **c–g,** Tyrosine phosphorylation of PDGFRs. Cells expressing PDGFR-α (**c**) or PDGFR-β (**d**), or CCD1070 cells (expressing PDGFR-α and PDGFR-β; **e, f**) were serum-starved and incubated in the presence or absence of PDGF-DD, PDGF-AA or PDGF-BB at the indicated concentrations for 10 min. Whole-cell lysates were prepared and analysed using a specific two-site ELISA for incorporation of phosphotyrosine into PDGFR-α (**c, e**) or PDGFR-β (**d, f**). In **e** and **f**, PDGFR-α/β heterodimers, as well as monoclonal-antibody-specific PDGFR homodimers contribute to the anti-phosphotyrosine signal after treatment with PDGF-BB and PDGF-DD, but not PDGF-AA. **g,** CCD1070 fibroblasts were treated with the indicated PDGFs at 200 ng ml^{-1} for 10 min. Lysates were then immunoprecipitated (IP) with antibodies against PDGFR-α or PDGFR-β and subjected to western blotting with anti-phosphotyrosine monoclonal antibody (anti-PY) or with antibodies against PDGFR-α or PDGFR-β as indicated. Blots were visualized using enhanced chemiluminescence. **h,** Inhibition of binding of ^{125}I-labelled PDGF-AA to CCD1070 fibroblasts. Binding was competed with PDGF-DD, PDGF-BB or PDGF-AA (see Methods; s.d. did not exceed 15%). **i,** PDGFR heterodimerization in cells expressing both PDGFRs. CCD1070 fibroblasts were serum-starved, incubated with the indicated PDGF for 10 min and lysed. Formation of PDGFR-α/β heterodimer complexes was assayed (see Methods).

PDGF-AA was unable to induce heterodimerization. Thus, PDGF-DD-induced tyrosine phosphorylation of PDGFR-α may be explained by the formation of PDGFR-α/β signalling complexes and concomitant phosphorylation.

A possible explanation for the existence of PDGF-DD is its potential for stimulating cells expressing either PDGFR-β or both PDGFRs, while not affecting nearby cell populations expressing only PDGFR-α. PDGF-DD does not compensate for the predominantly lethal phenotype observed in PDGF-knockout mice[1–5], indicating that it may have unique spatial and temporal functions. Cleavage of the PDGF-DD CUB domain represents a strategy by which growth-factor activity may be regulated by adjacent or distant cell populations dependent on protease expression, recognition and cleavage. Of particular interest is the high expression of PDGF-DD in the adrenal gland, an organ that secretes hormones particularly during stress[20]. It is possible that PDGF-DD has a systemic and/or neuroendocrine function during, for example, trauma or fibrotic disease. The lack of a sequence in PDGF-D for cell-surface retention supports this idea. In addition, we have observed increased PDGF-DD expression in cancer tissues and cell lines. Considering the role that PDGFs have in malignancy, it is possible that inappropriate expression of PDGF-DD contributes to certain cancers. We are currently generating the necessary reagents to validate further the role of PDGF-DD in human diseases, as well as exploring the possibility of exploiting its growth-promoting properties for therapeutic purposes.

Note added in proof: After acceptance of this manuscript, we learnt of the work of Bergsten *et al.* in this issue, which describes an identical PDGF-D sequence[21]. □

Methods

Cells and growth factors.

NIH 3T3 fibroblasts, CCD1070 fibroblasts, MG-63 osteosarcoma (ATCC) and 293-EBNA cells (Invitrogen) were obtained from commercial sources. CCD1070s contain ~50,000 PDGFR-α molecules and ~95,000 PDGFR-β molecules. 32DαR (a gift from J. Pierce, National Cancer Institute, Bethesda, Maryland) and HR5βR cells containing ~50,000 PDGFR-α molecules and ~55,000 PDGFR-β molecules, respectively, are described elsewhere[9,19].

Identification and isolation of human and murine PDGF-D cDNAs.

A technology (R. Shimkets, manuscript in preparation) that provides DNA-sequence information for the coding regions of expressed genes was used to identify a partial PDGF-D cDNA. Full-length

© 2001 Macmillan Magazines Ltd

PDGF-D cDNA (1,828 bp) was obtained by rapid amplification of cDNA ends (RACE). The sequence of PDGF-D encoding amino acids 24–370 was amplified from a human pituitary library (Clontech) by PCR using forward primer 5′-CTCGTCGAATTCACCCCGCAGAGCGCATCCATCAAAGC and reverse primer 5′-CTCGTCCTCGAGTCGAGGTGGTCTTG AGCTGCAGATACA. Murine PDGF-D was amplified from a murine brain library (Clontech) by PCR using forward primer 5′-CGCGGATCCATG-CAACGGCTCGTTTTAGTCTCCATTCTCC and reverse primer 5′- CGCGGATCCTTATCGAGGTG-GTCTTGAGCTGCAGATACAGTC.

Genomic organization of the PDGF D gene.

The genomic organization of human PDGF-D was deduced from hits (>99%) generated by a BLASTN analysis of the genomic clones with GenBank accession nos AC026640, AC023129, AC024052 and AC067870.

Purification of recombinant PDGF-DD.

PDGF-D was expressed in HEK 293 cells grown on porous microcarriers (Cultisphere-GL; Hyclone, Logan, Utah) in 1-litre spinner flasks. Cells were grown in DMEM/F12 media containing 1% penicillin/streptomycin in the presence or absence of 5% FBS. Conditioned medium was loaded onto a POROS HS50 column (PE Biosystems, Foster City, California) pre-equilibrated with 20 mM Tris–acetate, pH 7.0. After washing with equilibration buffer, bound proteins were eluted with a step NaCl gradient (0.25 M, 0.5 M, 1.0 M and 2.0 M). Fractions containing PDGF-DD p35 (1.0 M NaCl step elution) or p84 (0.5 M NaCl step elution) were pooled, dialysed and then loaded onto a POROS MC20 column precharged with nickel sulphate (PE Biosystems). Bound proteins were eluted with a linear gradient of imidazole (0–0.5 M). Protein purity was estimated to be >95% by SDS–PAGE (4–20% Tris–glycine gradient gel; Invitrogen) analysis.

Real-time quantitative PCR expression analysis.

RNA samples comprising normal human tissues were obtained commercially (Clontech; Invitrogen; Research Genetics). Real-time quantitative PCR was carried out using an ABI Prism 7700 sequence-detection system (PE Applied Biosystems) using *Taq*Man reagents (PE Applied Biosystems). RNAs were normalized using human β-actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) *Taq*Man probes according to the manufacturer's instructions. Equal quantities of normalized RNA were used as template in PCR reactions with PDGF-D-specific reagents to obtain threshold cycle (C_T) values. For graphical representation, C_T numbers were converted to per cent expression, relative to the sample exhibiting the highest level of expression. Primers used for PDGF-D analysis were as follows: forward, 5′-CGCTTGGCATCATCATTGAG; reverse, 5′-CGGTATCGAGGCAGGTCATAC; *Taq*Man probe, 5′-FAM-TCCAGGTCAACTTTTGACTTCCGGTCA-TAMRA.The FAM (6-carboxyfluorescein) reporter is covalently linked to the 5′ end of the probe; TAMRA (6-carboxy-*N*,*N*,*N*′,*N*′-tetramethyl-rhodamine), located at the 3′ end of the probe, is used for quenching. Primers used for PDGF-B analysis were as follows: forward, 5′-AAGATCGAGATTGTGCGGAAGA; reverse, 5′-ACTTGCAT-GCCAGGTGGTCT; *Taq*Man probe, 5′-FAM-CCAGCGTCACCGTGGCCTTCTTAA-TAMRA.

BrdU-incorporation assay.

Cells were cultured in 96-well plates to ~100% confluence, washed, fed with DMEM and starved for 24 h. Recombinant PDGF-DD, PDGF-AA or PDGF-BB was then added at the indicated concentration to cells for 18 h. The BrdU-incorporation assay was carried out according to the manufacturer's specifications (Roche) with an incorporation time of 5 h. In some experiments, BrdU incorporation induced by PDGF-AA, PDGF-BB or PDGF-DD was blocked by anti-hPDGFR-α (AF-307-NA) and/or anti-hPDGFR-β (AF385) neutralizing antibodies (R & D Systems, 10 μg ml^{-1}).

PDGFR-binding assays.

Adherent HR5βR, MG-63 and CCD1070 cells were resuspended in PBS with 5 mM EDTA,washed three times in binding medium (RPMI, 25 mM HEPES, pH 7.4, and 1 mg ml^{-1} BSA for HR5βR and 32DαR cells, or DMEM, 25 mM HEPES, pH 7.4, and 1 mg ml^{-1} BSA for CCD1070 cells). PDGF-AA labelled with ^{125}I using the Chloramine T method, or ^{125}I-labelled PDGF-BB (New England Nuclear) was added to 0.5×10^6 HR5βR or CCD1070 cells, or 1×10^6 32DαR cells in the presence of increasing concentrations of unlabelled ligand, and incubated on ice for 90 min. Bound and unbound ligand were separated using an oil-phase separation method and counted using a Beckman gamma counter[19].

PDGFR tyrosine-phosphorylation assays.

For the experiments shown in Fig. 3c–f, CCD1070 fibroblasts were grown to confluence, starved and then stimulated as described[19]. Tyrosine phosphorylation of receptors was quantified as described[19] using monoclonal antibodies alphaR10 and 1B5B11 (5 μg ml^{-1}) to capture PDGFR-α or PDGFR-β, respectively. Anti-phosphotyrosine antibody (2.5 μg ml^{-1}; Transduction Laboratories) was used to measure tyrosine phosphorylation of PDGFRs. For the experiments shown in Fig. 3g, CCD1070 fibroblasts were serum-starved for 18 h and then stimulated with 200 ng ml^{-1} PDGF-DD, PDGF-AA or PDGF-BB for 10 min. Whole-cell lysates were solubilized in RIPA buffer (25 mM Tris–HCl pH 7.5, 150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS, 5 mM EDTA, 10 mM sodium pyrophosphate, 50 mM sodium fluoride, 1 mM sodium orthovanadate, 1 mM phenylmethylsulphonyl fluoride, 10 μg ml^{-1} leupeptin, 10 μg ml^{-1} pepstatin and 1 μg ml^{-1} aprotinin)[9], sonicated, incubated with antibodies against PDGFR-α (SC-338) or PDGFR-β (SC-432; Santa Cruz Biotechnology; 5 μg each) and precipitated with protein G–agarose. SDS–PAGE sample buffer with 100 mM dithiothreitol was added, and samples were fractionated on 7.5% SDS–polyacrylamide gels. After electrophoretic transfer to Immobilon P membranes (Millipore), filters were blocked and then incubated with antibodies against PDGFR-α or PDGFR-β (1:500) or phosphotyrosine (1:1000), according to the manufacturer's instructions. Bound antibody was detected after incubation for 1 h with goat anti-rabbit immunoglobulin G (IgG; whole molecule; 1:2,000) or goat anti-mouse IgG (heavy and light chains; 1:10,000) conjugated to horseradish peroxidase (Boehringer). Enhanced chemiluminescence (Amersham) was carried out according to the manufacturer's instructions.

PDGFR-heterodimerization assay.

CCD1070 fibroblasts were grown to confluence, starved and then stimulated with 10 ng ml^{-1} PDGF-AA, 10 ng ml^{-1} PDGF-BB or 100 ng ml^{-1} PDGF-DD for 10 min at 37 °C. Lysates were prepared and heterodimeric PDGFR-α/β complexes were detected with a specific two-site ELISA using anti-PDGFR-β monoclonal antibody 1B5B11 (5 μg ml^{-1}) to capture PDGFR-β and anti-PDGFR-α antibody 3979 (2.5 μg ml^{-1}) to detect bound PDGFR-α as described[19]. The amount of PDGF-α/β heterodimers was quantified as described using a kinetic softmax program[19].

RECEIVED 14 OCTOBER 2000; REVISED 17 JANUARY 2001; ACCEPTED 26 JANUARY 2001; PUBLISHED 11 APRIL 2001.

1. Bostrom, H. *et al. Cell* 85, 863–873 (1996).
2. Leveen, P. *et al. Genes Dev.* 8, 1875–1887 (1994).
3. Lindahl, P., Johansson, B., Leveen, P. & Betsholtz, C. *Science* 277, 242–245 (1997).
4. Soriano, P. *Genes Dev.* 8, 1888–1896 (1994).
5. Soriano, P. *Development* 124, 2691–2700 (1997).
6. Heldin, C. & Westermark, B. *Physiol. Rev.* 79, 1283–1316 (1999).
7. Li, X., Ponten, A., Aase, K., Karlsson, L. & Abramsson, A. *Nature Cell Biol.* 2, 302–309 (2000).
8. Hamada, T., Ui-Tei, K. & Miyata, Y. *FEBS Lett.* 475, 97–102 (2000).
9. Matsui, T. *et al. Science* 243, 800–804 (1989).
10. Claesson-Welsh, L., Eriksson, A., Westermark, B. & Heldin, C. *Proc. Natl Acad. Sci. USA* 86, 4917–4921 (1989).
11. Bowen-Pope, D., Hart, C. & Seifert, R. *J. Biol. Chem.* 264, 2502–2508 (1989).
12. Nakai, K. & Kanehisa, M. *Genomics* 14, 897–911 (1992).
13. Sonnhammer, E., Eddy, S. & Durbin, R. *Proteins* 28, 405–420 (1997).
14. Bucher, P. & Bairoch, A. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 53–61 (1994).
15. Thielens, N., Bersch, B., Hernandez, J. & Arlaud, G. *Immunopharmacology* 42, 3–13 (1999).
16. Giese, N., Robbins, K. & Aaronson, S. *Science* 236, 1315–1318 (1987).
17. LaRochelle, W., May-Siroff, M., Robbins, K. & Aaronson, S. *Genes Dev.* 5, 1191–1199 (1991).
18. Heid, C., Stevens, J., Livak, K. & Williams, P. *Genome Res.* 6, 986–994 (1996).
19. Lokker, N. *et al. J. Biol. Chem.* 272, 33037–33044 (1997).
20. Pignatelli, D., Magalhaes, M. & Magalhaes, M. *Horm. Metab. Res.* 30, 464–474 (1998).
21. Bergsten, E. *et al. Nature Cell Biol.* 3, 512–516 (2001) (this issue).

ACKNOWLEDGEMENTS

We thank D. Andrew, S. Minskoff and B. Gould-Rothberg for discussions, and S. Colman for chromosome mapping.

Correspondence and requests for materials should be addressed to W.J.L. The nucleotide sequences of human and murine PDGF-D have been deposited at GenBank under accession nos AF335584 and AF335583, respectively.

© 2001 Macmillan Magazines Ltd

Attachment B

RESEARCH/

BY-PASSING IMMUNIZATION: BUILDING HIGH AFFINITY HUMAN ANTIBODIES BY CHAIN SHUFFLING

James D. Marks[1], Andrew D. Griffiths[1], Magnus Malmqvist[1], Tim P. Clackson[1], Jacqueline M. Bye[1,2] and Greg Winter[1,3,*]

[1]MRC Centre for Protein Engineering, MRC Centre, Cambridge CB2 2QH, U.K. [2]MRC Immunopathology Unit, Department of Immunology, AFRC Institute of Animal Physiology and Genetics, Babraham, Cambridge CB2 4AT, U.K. [3]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K. *Corresponding author.

Diverse antibody libraries can be displayed on the surface of filamentous bacteriophage, and selected by panning of the phage with antigen. This allows human antibodies to be made directly *in vitro* without prior immunization, thus mimicking the primary immune response[1]. Here we have improved the affinity of one such "primary" antibody by sequentially replacing the heavy and light chain variable (V) region genes with repertoires of V-genes (chain shuffling)[2] obtained from unimmunized donors. For a human phage antibody for the hapten 2-phenyloxazol-5-one (phOx) (K_d = 3.2 × 10^{-7} M), we shuffled the light chains and isolated an antibody with a 20 fold improved affinity. By shuffling the first two hypervariable loops of the heavy chain, we isolated an antibody with a further 15-fold improved affinity. The reshuffled antibody differed in five of the six hypervariable loops from the original antibody and the affinity for phOx (K_d = 1.1 × 10^{-9} M) was comparable to that of mouse hybridomas from the tertiary immune response. Reshuffling offers an alternative to random point mutation for affinity maturation of human antibodies *in vitro*.

For serotherapy, monoclonal antibodies would ideally be of human origin, but human hybridomas are difficult to make and require human immunization (see ref. 3 for review). New technologies have prompted new solutions. For example, gene technology has prompted the 'humanizing' of rodent antibodies by transplanting their hypervariable loops into a human antibody[4-8], leading to clinical application[9]. The use of the polymerase chain reaction[10] (PCR), to clone and express antibody V-genes[11,12] and phage display technology[13,14] to select antibody genes encoding fragments with binding activities[15] has resulted in the isolation of antibody fragments from repertoires of PCR amplified V-genes using immunized mice or humans[2,16] thus by-passing conventional hybridoma technology.

Recently, we reported the isolation of human antibody fragments directed against both small (hapten) and large (protein) antigens from the same single chain Fv (scFv)[17,18] library (3 × 10^7 members) made from the V-genes of unimmunized healthy blood donors and displayed on the surface of bacteriophage[1]. The process by-passes immunization by mimicking immune selection. Indeed, the antibody fragments were highly specific and had affinities typical of a primary immune response (K_d = 1 – 5 × 10^{-7} M). The technology appears to have the potential to make human antibodies entirely *in vitro*, but for most practical applications the antibodies need higher affinities typical of later immune responses.

Affinity maturation can be mimicked *in vitro* by making point mutations in the V-genes, for example by using an error-prone polymerase, and selecting mutants for improved affinity[19]. Alternatively, new combinations of antibody heavy and light chains can be made by recombining a single heavy or light chain with a library of partner chains (chain shuffling). Chain shuffling has been used to make new combinations of heavy and light chains with hapten binding activities from the V-genes of immunized animals but affinities of the shuffled antibodies were not measured[2,20]. An attempt to derive hapten binding antibodies by reshuffling the V-genes from an immunized source with those from a naive source failed, prompting the authors to assert that "redesign of antibodies through recombination of a somatically mutated chain with a naive partner may be a difficult process"[20].

For this work, we started with the human antibody (αphOx-15) directed against the hapten 2-phenyloxazol-5-one (phOx) that had been isolated from a phage display library made from unimmunized human donors[1]. Both heavy and light chains of α-phOx-15 are somatically mutated. Using repertoires of heavy and light chain V-genes from unimmunized donors, we reshuffled the heavy chain with the repertoire of light chains, and vice-versa to make shuffled somatically mutated antibodies with higher affinities.

RESULTS

Light chain shuffling. A scFv fragment (αphOx-15) directed against the hapten phOx was isolated from a phage antibody library constructed from the heavy (VH) and light (Vκ and Vλ) chain genes from the peripheral blood lymphocytes of unimmunized human donors[1]. The VH gene of αphOx-15 was assembled with a repertoire of Vκ and Vλ genes from the same unimmunized donors to make shuffled scFv genes[2], and cloned into the phagemid vector pHEN1 (ref. 21) for display as a fusion with gene 3 coat protein[15]. After transformation, the pha-

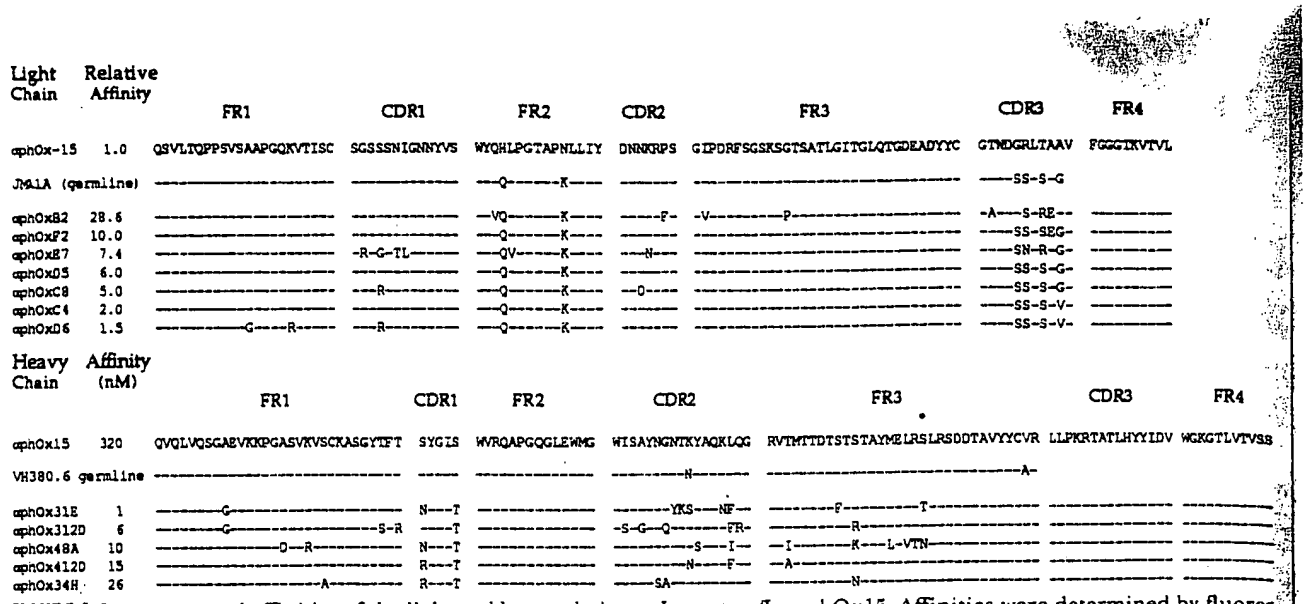


FIGURE 1 Sequences and affinities of the light and heavy chains of phOx binders from the shuffled libraries. The sequences of the light chains are compared to αphOx-15 and the most homologous Vλ germ-line gene, JMλ1A (ref. 1). The sequences of the heavy chains are compared to αphOx-15 and the most homologous germline gene VH380.6 (ref. 1). Relative affinities were determined by inhibition ELISA and are expressed as I_{50} mutant/I_{50} αphOx15. Affinities were determined by fluorescence quench titration. All antibodies bound phOx specifically (did not bind BSA in an ELISA and binding to phOx-BSA coated microtitre plates could be inhibited by soluble phOx-GABA). *Location of the cloning site for the heavy chain repertoire.

gemid library (2×10^6 clones) appeared diverse by BstNI fingerprinting[2], was rescued with helper phage[1] and subjected to panning on phOx-BSA coated tubes[1]. For expression of soluble scFv, the phage eluted from the tubes were used to infect a non-suppressor strain of bacteria[21] (for details see Experimental Protocol).

To identify clones with improved affinities, the binding of soluble scFvs to phOx-BSA were compared by ELISA. After a single round of panning, soluble scFv from 59/192 clones bound to phOx-BSA with a stronger signal than αphOx-15 scFv whereas before panning, none of 192 clones gave a stronger signal. Six of these clones, and a further 4 clones from a second round of panning, were sequenced. Six unique Vλ light chains were found, all from the same Vλ1-gene family and probably the same germ-line gene as the αphOx-15 light chain (Fig. 1). The human Vλ chains were mutated at a range of sites, diverging by 0 to 9 amino acid residues from the putative Vλ germ-line gene (VλJM1A). The clustering of residue changes, particularly in CDR3, indicates that the mutant light chains were derived directly from the germ-line Vλ-gene rather than the αphOx-15 light chain (Fig. 1).

The scFv fragments from eight different clones were ranked by competition for binding to phOx-BSA with soluble phOx hapten[22], and the "relative affinities" were found to be up to 27 fold higher than αphOx-15 (Fig. 1). The affinities of αphOx-15 and αphOxB2 (the clone with the highest relative affinity) were also measured directly by fluorescence quench titration. The affinity of αphOxB2 was found to be 1.5×10^{-8} M (20 fold higher than αphOx-15) (Table 1). The kinetics of binding (off-rates) of purified αphOx-15 and αphOxB2 scFvs to phOx modified BSA were determined by real-time biospecific interaction analysis based on surface plasmon resonance (SPR, Pharmacia BIAcore)[23,24]. The off-rate was much slower for αphOxB2 but calculated on-rates (k_{off}/K_d) were similar (Table 1). Thus the improved affinity of αphOxB2 is due to its slower off-rate.

Heavy chain shuffling. The reshuffled heavy chain library was prepared as described in the Experimental Protocol. Briefly, a repertoire of VH genes (VH1 family) was amplified by PCR from the IgG and IgM mRNAs of unimmunized donors using primers based in the first and third framework regions. The VH repertoire, which encodes the first two hypervariable loops and three framework regions, was cloned into a vector encoding the third hypervariable loop and the light chain of α-phOxB2. The resulting library (2×10^5 clones) was panned on phOx and soluble scFv screened by ELISA for binding after each round of selection.

TABLE 1 Affinities and kinetics of binding to phOx of original isolate (αphOx15) and chain shuffled mutants.

Clone	Residue changes (from αphOx-15)	K_d*(M)	k_{off}†(s^{-1})	k_{on}§(M^{-1} s^{-1})
Original Isolate				
αphOx-15	0	$3.2\kappa0.1 \times 10^{-7}$	$4.3\kappa0.6 \times 10^{-1}$	1.3×10^6
New Light chain				
αphOxB2	10	$1.5\kappa0.6 \times 10^{-8}$	$1.7\kappa0.4 \times 10^{-2}$	1.1×10^6
New Light chain and heavy chains				
αphOx34H	16	$2.6\kappa0.7 \times 10^{-8}$	$7.3\kappa0.8 \times 10^{-3}$	2.8×10^5
αphOx412D	15	$1.5\kappa0.4 \times 10^{-8}$	$5.8\kappa0.6 \times 10^{-3}$	3.9×10^5
αphOx48A	22	$1.0\kappa0.2 \times 10^{-8}$	$2.5\kappa0.2 \times 10^{-3}$	2.5×10^5
αphOx312D	20	$6.0\kappa1.1 \times 10^{-9}$	$3.5\kappa0.6 \times 10^{-3}$	5.8×10^5
αphOx31E	20	$1.1\kappa0.4 \times 10^{-9}$	$3.8\kappa0.5 \times 10^{-3}$	3.5×10^6

*Measured by fluorescence quench titration. †Measured by surface plasmon resonance in BIAcore (Pharmacia). §Calculated from k_{off}/K_d.

Before selection 0/94 clones bound to phOx whereas after 3 and 4 rounds of selection, 38/94 and 51/94 clones bound to phOx. Supernatants from all 90 clones were screened by SPR for dissociation from phOx-BSA. All 90 clones had slower off-rates than αphOxB2. These clones were grouped according to off-rate and BstNI restriction pattern and eight clones were sequenced (Fig. 1) revealing 5 unique sequences. All 5 were derived from the same germline VH-gene (VH380.6, ref. 1) as αphOx-15 and αphOxB2 but had an additional 5 to 12 residue changes (Fig. 1 and Table 1). Residue 35 was changed from serine to threonine in all 5 mutants.

The affinities of three of the mutants were shown by fluorescence quench titration to be greater than αphOxB2 (Table 1). The affinities ranged from 2.6×10^{-8} M to 1.1×10^{-9} M (12 to 320-fold higher than αphOx-15 and 0.6 to 15-fold higher than αphOxB2). All five mutants had slower off-rates than αphOx-15 or αphOxB2. The highest affinity antibody, αphOx31E, had a faster on-rate than αphOx-15 or αphOxB2.

DISCUSSION

Previously, we used phage display and the V-genes from unimmunized donors to make antibody fragments against both small (hapten) and large (protein) antigens with affinities typical of the primary immune response. While the approach is potentially useful for making therapeutic human antibodies, we need to find ways of increasing the antibody affinities. Here we have shown that this can be accomplished by chain shuffling. We diversified the structure of an antibody by first shuffling light chains, then heavy chains, while retaining the third hypervariable loop of the heavy chain. Much of the sequence and structural variation of antigen binding sites is encoded by this loop, which is located at the center of the antigen binding site[25]. By retaining it, while shuffling the other loops, we aimed to diversify the structure without disrupting the key features of the antigen binding site.

We chose the hapten phOx for our model experiments, as the immune response and affinity and kinetic maturation is well studied[26-28]. The affinities of αphOxB2, from the light chain shuffled library, and the 5 mutants from the heavy chain shuffled library are comparable to that of mouse hybridomas from the secondary or tertiary immune response to the same hapten[26] (Fig. 2). Indeed, of anti-phOx hybridomas from the mouse secondary or tertiary response, only 2 of 24 had a higher affinity than αphOx31E (ref. 26).

The improvement in affinity results almost exclusively from a slower off-rate. Somatic hypermutation of the V-genes used in the murine primary immune response to phOx also improves affinity mainly by slowing the off-rate[26]. The results suggest that our washing and binding conditions favor the selection of phages with slower off-rates rather than faster on-rates in contrast to the suggestion of Garrard et. al.[29]. As we build antibodies with higher and higher affinities, it becomes increasingly likely that the best binders will remain attached to the solid phase, necessitating more vigorous elution conditions.

In vivo, affinity maturation occurs by random mutation of the original heavy and light chain pairings and by the appearance of new heavy and/or light chain pairings (repertoire shift)[27,28]. We can simultaneously mimic aspects of both processes *in vitro* by tapping the natural pool of diverse unmutated and mutated heavy and light chains via chain shuffling. Using V-genes derived from an immunized mouse, we had previously shown that new partners could arise from different V-gene families[2]. In the present study, both the light chains and heavy chains are derived from the same germline gene and the antibodies differ only by point mutations. Nevertheless the repertoire of mutants should differ from those generated by random mutation of the αphOx-15 antibody in two respects. Firstly, the V-genes encoding the shuffled chains have been selected from the mRNA of B-lymphocytes and are more likely to be functional. In contrast, *in vitro* random mutagenesis, for example using an error prone polymerase[19], is likely to result in many mutants that would compromise chain folding, particularly if multiple mutations were introduced into the same gene. Secondly, with *in vitro* mutagenesis mutations are introduced directly into αphOx-15 whereas with chain swapping, mutations are introduced into the corresponding germline genes. This could allow any deleterious mutations in αphOx-15 to be replaced more readily.

A shuffling strategy may be applicable to protein antigens as well as haptens. Although there are a larger number of contacts between protein and antibody, and the chances of disrupting multiple favorable contacts by shuffling is greater, this may be compensated by the loss of multiple unfavorable contacts.

One advantage of building an artificial immune system is that by allowing heavy chains to sample other light chains, and vice-versa we employ a strategy that is not open to the immune system. Thus shuffling enlarges the repertoire size, enhancing the chances of finding higher affinity antibodies[30], and in principle allowing chains with deleterious mutations[31] to be replaced by others. Shuffling chains and hypervariable loops appears to be a powerful way of diversifying antibody structure, and the pool of rearranged V-genes from unimmunized donors provides a rich source of genetic diversity.

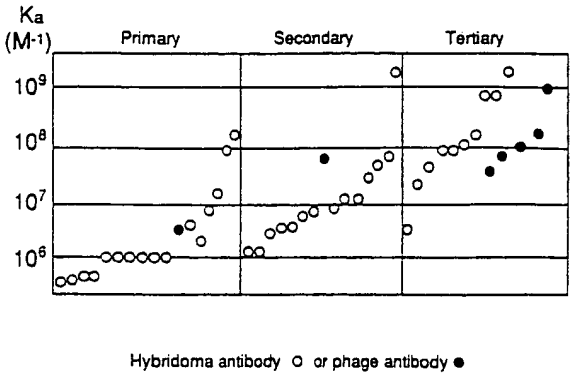


FIGURE 2 Comparison of affinities of anti phOx antibodies from hybridomas and from phage antibodies. Affinity constants (Ka) for anti-phOx hybridomas from primary, secondary and tertiary responses from immunized mice (data taken from ref. 26) are compared with data (Table 1) for phage antibodies from naive phage library (primary), light chain shuffled (secondary) and heavy chain shuffled (tertiary) libraries.

EXPERIMENTAL PROTOCOL

Construction of a reshuffled light chain library. A scFv library was assembled[2] from the VH gene of αphOx-15 and a Vλ and Vκ repertoire[1] using PCR. To avoid contamination with the original light chain, the VH gene of αphOx-15 was subcloned into the vector pJM-1 (ref. 2), amplified by PCR using primers HuVH1aBACK and HuJH6FOR[1], purified on a 2% (w/v) agarose gel and isolated from the gel using Geneclean (Bio-101). Reshuffled scFv repertoires were PCR assembled[1] from the phOx-15VH DNA, linker DNA and the same human Vλ and Vκ gene repertoires used to construct the primary library from

which αphOx-15 was isolated[1]. The repertoires were digested with NcoI and NotI, purified on a 1.5% (w/v) agarose gel, electroeluted[32], precipitated with ethanol and ligated into the vector pHEN-1 (ref. 21) digested with NcoI and NotI. The ligation mix was used to transform electrocompetent[33] *E. coli* TG1 (ref. 34). Cells were grown for 1 hour in 1 ml of SOC[32] and then plated on TYE[35] medium with 100 μg/ml ampicillin 1% (w/v) glucose. Colonies were scraped off the plates into 5 ml of 2 × TY[35] broth containing 100 μg/ml ampicillin, 1% (w/v) glucose and 15% glycerol.

Construction of a reshuffled heavy chain library. A scFv library was prepared containing the VH CDR3 and Vλ of αphOxB2 and a repertoire of human VH1 genes. To eliminate potential contamination with the original heavy chain, the human VH1 pseudogene DP-22 (ref. 36) was amplified using PCR from an M13 template using the primers HuVH1BACK-SFI (ref. 1) and HuVH1FR3FOR (5'-GGC CGT G/CTC AGA TCT CAG-3'), digested with NcoI and BglII, gel purified and ligated into the vector pHEN-1phOxB2 digested with NcoI and BglII. The resulting vector, pHEN-1-ΨVHB2, contained the DP-22 VH1 pseudogene and the VH CDR3 and Vλ of αphOxB2. To prepare a repertoire of human VH1 genes, human PBL RNA was primed in separate reactions with HuIgG1-4CH1FOR and HuIgMFOR and 1st strand cDNA synthesized[1]. The first strand cDNA was used as a template for PCR amplification as previously described[1] using the primers HuVH1aBACK and HuVH1FR3FOR. Restriction sites were appended to the repertoires by reamplification using the primers HuVH1BACKSFI and HuVH1FR3FOR. The VH1 repertoires were digested with NcoI and BglII, purified on a 1.5% (w/v) agarose gel, electroeluted, precipitated with ethanol and ligated into the vector pHEN-1-ΨVHB2 digested with NcoI and BglII. The ligation mix was used to transform electrocompetent *E. coli* TG1. Cells were grown for 1 hour in 1 ml of SOC and then plated on TYE medium with 100 μg/ml ampicillin and 1% (w/v) glucose. Colonies were scraped off the plates into 5 ml of 2 × TY broth containing 100 μg/ml ampicillin, 1% (w/v) glucose and 15% glycerol.

Selection of reshuffled libraries. To rescue phagemid particles, 50 ml of 2 × TY containing 100 μg/ml ampicillin and 1% (w/v) glucose (2 × TY AMP-GLU) were inoculated with 10^9 bacterial cells from the library glycerol stock, grown with shaking at 37°C to an A_{590} of 0.9 and then 5 ml added to 50 ml of 2 × TY AMP-GLU prewarmed to 37°C. 2 × 10^{10} plaque forming units of VCS-M13 (Stratagene) were added and the mixture incubated at 37°C without shaking for 1 hour. The mixture was then added to 500 ml of 2 × TY broth containing 100 μg ampicillin/ml and 25 μg kanamycin/ml and grown overnight at 37°C with shaking. Phage particles were purified and concentrated as previously described[1]. Two rounds (reshuffled light chain library) or four rounds (reshuffled heavy chain library) of enrichment for phOx binding phage were performed in phOx-BSA coated immunotubes (Nunc) (10 μg/ml of 14ox/BSA for selection of the reshuffled light chain library and 10 μg/ml of lox/BSA for selection of the reshuffled heavy chain library). After each round of enrichment, *E. coli* TG1 were reinfected with eluted phage and rescued to provide phage for the next round of panning. For soluble scFv expression, eluted phage was used to infect *E. coli* HB2151 (ref. 37).

Initial characterization of binders with new light chains. Soluble scFv was induced[38] from 94 colonies from each round of selection and analyzed for binding to phOx by ELISA[1]. Twelve clones with ELISA signals stronger than αphOx-15 were sequenced[39] revealing 8 unique clones. The relative affinities of these 8 clones were determined by inhibition ELISAs. For inhibition ELISAs[22], microtiter wells were coated overnight with 100 μg/ml phOx-BSA in PBS and blocked for 2 hours at 37°C with 2% milk powder in PBS. Dilutions of scFv previously determined to result in significant reduction of ELISA values after two-fold dilution were mixed with phOx (10^{-3}–10^{-7} M) in the wells and incubated for 1.5 hours at RT. Bound soluble scFv was detected by ELISA[1]. The concentration of phOx resulting in a 50% reduction in ELISA signal (I_{50}) was calculated for each mutant and compared to that obtained for αphOx15 to determine the relative affinity. Relative affinities, but not the I_{50} value (6.0–400 μM), correlated with affinities measured by fluorescence quench (Fig. 1 and Table 1). Affinities and off-rates of the clone with the highest relative affinity (αphOxB2) as well as αphOx-15 were determined as described below.

Initial characterization of binders with new heavy chains. Soluble scFv was induced[38] from 94 colonies from each round of selection and analyzed for binding to phOx by ELISA[1]. The off-rates of soluble scFv from all ninety positive clones from the third and fourth round of selection were determined by BIAcore (see below) and the clones then grouped according to off-rate and BstNI fingerprint[1]. Eight representative clones were sequenced[39] revealing 5 unique clones. Affinities and off-rates of these 5 clones were determined as described below.

Affinity measurements. Two liter cultures of *E. coli* HB2151 harboring the appropriate phagemid were induced[38] and the soluble scFv affinity purified[2] from the supernatant using the C-terminal peptide tag[40]. For affinity determinations, fluorescence quench titration with the hapten 4-γ-amino-butyric acid methylene 2-phenyl-oxazol-5-one (phOx-GABA) was performed as described[26]. The affinity of αphOx-15 was determined[26,41] using a regime of hapten excess as described previously[1]. Data were averaged from 3 runs. For determination of the affinity of αphOxB2 and the 5 mutants from the shuffled heavy chain library, 100 nM scFv (a concentration ten times the preliminary estimate of the dissociation constant) was titrated with hapten and the fluorescence determined 1 min after each addition[26]. Excitation was at 280 nm and emission was monitored at 340 nm. Data were averaged from 3 to 5 runs. k_{off} was measured by real-time biospecific interaction analysis based on surface plasmon resonance (BIAcore, Pharmacia Biosensor AB)[23,24]. Affinity purified scFv proteins were fractionated on a calibrated FPLC Superdex 75 column (Pharmacia) to eliminate aggregates and the monomeric fraction then used for kinetic measurements. In a BIAcore flow cell, 1300 resonance units (RU) of 100 μg/ml phOx modified BSA (14 phOx/BSA) in 10 mM acetate buffer pH 4.0 was coupled to a CM5 sensor chip[42]. In another flow cell, the sensor chip was activated without phOx-BSA as a control. Adsorption and dissociation of αphOx15 (0.4 μM–2.3 μM) and the other scFvs (80 nM–400 nM) in PBS, 0.2 mM EDTA were measured under a constant flow of 6 μl/min. k_{off} was determined for αphOxB2 and the heavy chain shuffled mutants from the dissociation part of the sensorgram and for αphOx15 from the association part of the sensorgram[43] (necessitated by its rapid k_{off}).

Acknowledgments

J.D.M. was supported by the MRC AIDS Directed Programme, T.P.C. by the Louis Jeantet Foundation (Geneva), A.D.G. by the Cancer Research Campaign, M. M. by Pharmacia Biosensor AB and J.M.B. by a grant from the International Blood Group Reference Laboratory to Nevin Hughes-Jones. We thank Jefferson Foote for phOx-GABA, Cara Marks for assistance with protein purification and Andreas Matouschek for help with analysis of kinetics.

Received 25 February 1992; accepted 4 May 1992.

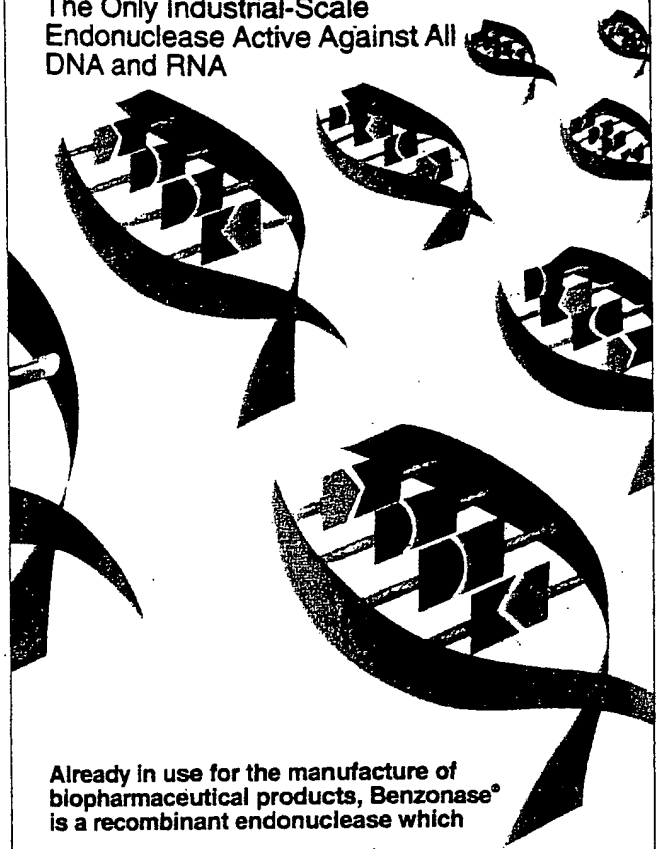
References

1. Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D. and Winter, G. 1991. By-passing immunization: Human antibodies from V-gene libraries displayed on phage. J. Mol. Biol. **222**:581–597.
2. Clackson, T., Hoogenboom, H. R., Griffiths, A. D. and Winter, G. 1991. Making antibody fragments using phage display libraries. Nature **352**:624–628.
3. Winter, G. and Milstein, C. 1991. Man-made antibodies. Nature **349**: 293–299.
4. Jones, P., Dear, P., Foote, J., Neuberger, M. and Winter, G. 1986. Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature **321**:522–525.
5. Riechmann, L., Clark, M., Waldmann, H. and Winter, G. 1988. Reshaping human antibodies for therapy. Nature **332**:323–327.
6. Queen, C., Schneider, W. P., Selick, H. E., Payne, P. W., Landolfi, N. F., Duncan, J. F., Avdalovic, N. M., Levitt, M., Junghans, R. P. and Waldmann, T. A. 1989. A humanized antibody that binds to the interleukin 2 receptor. Proc. Natl. Acad. Sci. USA **86**:10029–10033.
7. Tempest, P. R., Bremner, P., Lambert, M., Taylor, G., Furze, J. M., Carr, F. J., Harris, W. J. 1991. Reshaping a monoclonal antibody to inhibit human respiratory syncytial virus infection *in vivo*. Bio/Technology **9**:266–272.
8. Gorman, S. D., Clark, M. R., Routledge, E. G., Cobbold, S. P. and Waldmann, H. 1991. Reshaping a therapeutic CD4 antibody. Proc. Natl. Acad. Sci. USA **88**:4181–4185.
9. Hale, G., Dyer, M. J., Clark, M. R., Phillips, J. M., Marcus, R., Riechmann, L., Winter, G. and Waldmann, H. 1988. Remission induction in non-Hodgkin lymphoma with reshaped human monoclonal antibody CAMPATH-1H. Lancet **2**:1394–1399.
10. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**:1350–1354.
11. Orlandi, R., Gussow, D. H., Jones, P. T. and Winter, G. 1989. Cloning immunoglobulin variable domains for expression by the polymerase chain reaction. Proc. Natl. Acad. Sci. USA **86**:3833–3837.
12. Huse, W. D., Sastry, L., Iverson, S. A., Kang, A. S., Alting, M. M., Burton, D.

R., Benkovic, S. J. and Lerner, R. A. 1989. Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. Science **246**:1275–1281.
13. Smith, G. P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science **228**: 1315–1317.
14. Parmley, S. F. and Smith, G. P. 1988. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. Gene **73**:305–318.
15. McCafferty, J., Griffiths, A. D., Winter, G. and Chiswell, D. J. 1990. Phage antibodies: filamentous phage displaying antibody variable domains. Nature **348**:552–554.
16. Burton, D. R., Barbas, C. F., Persson, M. A. A., Chanock, R. M. and Lerner, R. A. 1991. A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic individuals. Proc. Natl. Acad. Sci. USA **88**:10134–10137.
17. Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S. M., Lee, T., Pope, S. H., Riordan, G. S. and Whitlow, M. 1988. Single-chain antigen-binding proteins. Science **242**:423–426.
18. Huston, J. S., Levinson, D., Mudgett, H. M., Tai, M. S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R. and Oppermann, H. 1988. Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85**:5879–5883.
19. Hawkins, R. E., Russell, S. J. and Winter, G. 1992. Selection of phage antibodies by binding affinity: mimicking affinity maturation. J. Mol. Biol. *In press*.
20. Kang, A. S., Jones, T. M. and Burton, D. R. 1991. Antibody redesign by chain shuffling from random combinatorial immunoglobulin libraries. Proc. Natl. Acad. Sci. USA **88**:11120–11123.
21. Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., Hudson, P. and Winter, G. 1991. Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. Nucl. Acids Res. **19**:4133–4137.
22. Rath, S., Stanley, C. M. and Steward, M. W. 1988. An inhibition enzyme immunosassay for estimating relative antibody affinity and affinity heterogeneity. J. Immunol. Methods **106**:245–249.
23. Jönsson, U., Fägerstam, L., Ivarsson, B., Lundh, K., Löfås, S., Persson, B., Roos, H., Rönnberg, I., Sjölander, S., Stenberg, E., Ståhlberg, R., Urbaniczky, C., Östlin, H. and Malmqvist, M. 1991. Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. BioTechniques **11**:620–627.
24. Jönsson, U. and Malmqvist, M. 1992. Real time biospecific interaction, p. 291–336. *In:* Advances in Biosensors. Turner (Ed.) JAI Press Ltd. San Diego, CA.
25. Chothia, C. and Lesk, A. M. 1987. Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. **196**:901–917.
26. Foote, J. and Milstein, C. 1991. Kinetic maturation of an immune response. Nature **352**:530–532.
27. Berek, C., Griffiths, G. M. and Milstein, C. 1985. Molecular events during maturation of the immune response to oxazolone. Nature **316**:412–418.
28. Berek, C. and Milstein, C. 1987. Mutation drift and repertoire shift in the maturation of the immune response. Immunol. Rev. **96**:23–41.
29. Garrard, L. J., Yang, M., O'Connell, M. P., Kelley, R. F. and Henner, D. J. 1991. Fab assembly and enrichment in a monovalent phage display system. Bio/Technology **9**:1373–1377.
30. Perelson, A. S. 1989. Immune network theory. Immunol. Rev. **110**:5–36.
31. Sharon, J. 1990. Structural correlates of high antibody affinity. Three engineered amino acids substitutions can increase the affinity of an anti-*p*-azophenylarsonate antibody 200-fold. Proc. Natl. Acad. Sci. USA **87**:4814–4817.
32. Sambrook, J. Fritsch, E. F. and Maniatis, T. 1990. Molecular Cloning—A Laboratory Manual. Cold Spring Harbor Laboratory, New York.
33. Dower, W. J., Miller, J. F. and Ragsdale, C. W. 1988. High efficiency transformation of *E. coli* by high voltage electroporation. Nucl. Acids Res. **16**:6127–6145.
34. Gibson, T. J. 1984. Studies on the Epstein-Barr virus genome. Ph.D. Thesis, University of Cambride, UK.
35. Miller, J. H. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, New York.
36. Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. and Winter, G. 1992. The repertoire of human germline VH sequences reveals fifty groups of VH segments with different hypervariable loops. J. Mol. Biol. *In press*.
37. Carter, P., Bedouelle, H. and Winter, G. 1985. Improved oligonucleotide site-directed mutagenesis using M13 vectors. Nucl. Acids Res. **13**: 4431–4443.
38. De Bellis, D. and Schwartz, I. 1990. Regulated expression of foreign genes fused to *lac*: control by glucose levels in growth medium. Nucl. Acids Res. **18**:1311.
39. Sanger, F., Nicklen, S. and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
40. Munro, S. and Pelham, H. R. B. 1986. An Hsp-like protein in the ER: Identity with the 78kd glucose regulated protein and immunoglobulin heavy chain binding protein. Cell **46**:291–300.
41. Eisen, H. N. 1964. Determination of antibody affinity for haptens and antigens by means of fluorescence quenching. Meth. Med. Research **10**:115–121.
42. Johnsson, B., Löfås, S. and Lindqvist, G. 1991. Immobilization of proteins to a carboxymethyldextran modified gold surface for BIAcore in surface plasmon resonance. Anal. Biochem. **198**:268–277.
43. Karlsson, R., Michaelsson, A. and Mattsson, L. 1991. Kinetic analysis of monoclonal antibody-antigen interactions with a new biosensor based analytical system. J. Immunol. Methods **145**:229–240.

Benzonase®

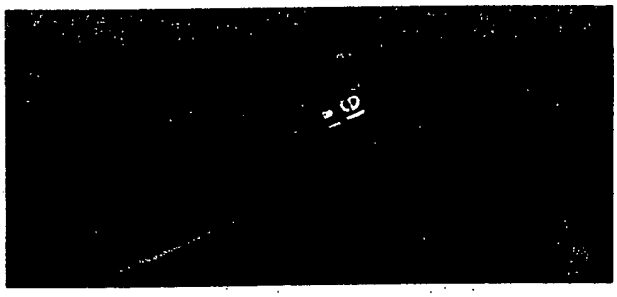
The Only Industrial-Scale Endonuclease Active Against All DNA and RNA



Already in use for the manufacture of biopharmaceutical products, Benzonase® is a recombinant endonuclease which

- Reduces DNA and RNA content to FDA standards
- Increases protein yield by reducing viscosity caused by nucleic acids
- Enables purification of cell-derived particles associated with nucleic acids

Benzonase® is a very stable enzyme, possessing an exceptionally high specific activity and is free of proteases. Benzonase® is supplied as a >90% or >99% pure product in glycerol.



The ultimate choice in down-stream processing

Available for large-scale use from

NYCOMED PHARMA

Helseholmen 1
DK-2650 Hvidovre
Denmark

Phone: +45 36 77 22 55
Fax: + 45 36 77 29 55

Benzonase® is a registered trademark of Benzon Pharma A/S, a company in the HAFSLUND NYCOMED corporation

Write in No. 251 on Reader Service Card

Attachment C

These data suggest that PntP2 may be a direct target for Rl/MAP kinase, as has been shown for ETS proteins in vertebrates[9,18]. Indeed, the P2 protein contains a single MAP kinase phosphorylation site (PLTP) motif in the Pointed domain corresponding to the consensus for MAP kinase phosphorylation[19] and can be phosphorylated by Rl/MAP kinase *in vitro* (Fig. 3*a*, lane 2). Phosphorylation is dependent on this single site because a mutant protein ($PntP2^{T151A}$) in which the threonine in the PLTP motif has been replaced by an alanine cannot be phosphorylated *in vitro* (Fig. 3*a*, lane 4). To test whether this mutation affects the function of PntP2 *in vivo*, we generated transgenic lines expressing the mutant transgene under the control of the *sev* enhancer (*sE-pnt*^{T151A}). Not only is this construct unable to rescue the *pnt* phenotype, rather it enhances the mutant phenotype (Fig. 2*g*). Even in a wild-type background the mutant protein prevents neural development of the R7 precursor (Fig. 1*i*) which results in the absence of R7 cells in the adult (Fig. 2*h*). This suggests that the mutant protein competes directly or indirectly with the wild-type protein. Similar to Pnt^{T151A}, a mutant form of vertebrate Elk-1 in which multiple MAP kinase phosphorylation sites have been deleted prevents serum response element-dependent transcription in a dominant-negative fashion[9].

Loss-of-function mutations in *yan*, which encodes another ETS domain protein, result in the recruitment of many R7 cells even in the absence of *sev* function[11] (Fig. 2*i*). In *yan*, *pnt* double mutants most of the ommatidia lack R7 and some outer photoreceptors (Fig. 2*k*), suggesting that the development of R7 cells in the absence of *yan* function depends on *pntP2* function. Although overexpression of *pntP2* under the control of the *sev* enhancer is not sufficient to transform cone cells into R7 cells in the wild type (Fig. 2*l*), many R7 cells form in a heterozygous *yan* background (Fig. 2*m*). These observations suggest that the mechanisms controlling R7 determination are sensitive to changes in the relative amounts of PntP2 as well as of Yan. The Yan protein contains 10 consensus sites for MAP kinase phosphorylation[19] and, like PntP2, can be phosphorylated by MAP kinase *in vitro* (Fig. 3*a*, lane 6).

Our data thus suggest a model in which MAP kinase activity induces neuronal differentiation by simultaneously inhibiting the Yan repressor and stimulating the PntP2 activator (Fig. 3*b*). The fact that both proteins contain an ETS DNA-binding domain known to recognize conserved target sequences[20] raises the possibility that these two proteins compete directly for binding sites present in the regulatory regions of target genes. □

Received 31 May; accepted 11 July 1994.

1. Greenwald, I. & Rubin, G. M. *Cell* **00**, 271–281 (1992).
2. Dickson, B. & Hafen, E. in *The Development of Drosophila* (eds Martinez-Arias, A. & Bate, M.) (Cold Spring Harbor Press, New York, 1993).
3. Dickson, B. & Hafen, E. *Curr. Opin. genet. Dev.* **4**, 64–70 (1994).
4. Brunner, D. *et al. Cell* **70**, 875–888 (1994).
5. Biggs, W. H. III *et al. EMBO J.* **13**, 1628–1635 (1994).
6. Chen, R. H., Sarnecki, C. & Blenis, J. *Molec. cell. Biol.* **12**, 915–927 (1992).
7. Lenormand, P. *et al. J. Cell Biol.* **122**, 1079–1088 (1993).
8. Treisman, R. *Curr. Opin. genet. Dev.* **4**, 96–101 (1994).
9. Marais, R., Wynne, J. & Treisman, R. *Cell* **73**, 381–393 (1993).
10. Chen, T., Bunting, M., Karim, F. D. & Thummel, C. S. *Devl Biol.* **151**, 176–191 (1992).
11. Lai, Z.-C. & Rubin, G. M. *Cell* **70**, 609–620 (1992).
12. Klämbt, C. *Development* **117**, 163–176 (1993).
13. Basler, K. & Hafen, E. *Development* **107**, 723–731 (1989).
14. Simon, M. A., Bowtell, D., Dodson, G. S., Laverty, T. R. & Rubin, G. M. *Cell* **07**, 701–716 (1991).
15. Rogge, R. D., Karlovich, C. A. & Banerjee, U. *Cell* **04**, 39–48 (1991).
16. Dickson, B., Sprenger, F., Morrison, D. & Hafen, E. *Nature* **360**, 600–603 (1992).
17. Basler, K., Siegrist, P. & Hafen, E. *EMBO J.* **0**, 2381–2386 (1989).
18. Janknecht, R., Ernst, W. H., Pingoud, V. & Nordheim, A. *EMBO J.* **12**, 5097–5104 (1993).
19. Marshall, C. J. *Curr. Opin. genet. Dev.* **4**, 82–89 (1994).
20. Wasylyk, B., Hahn, S. L. & Giovane, A. *Eur. J. Biochem.* **211**, 7–18 (1993).
21. Jürgens, G. E. W., Nüsslein-Volhard, C. & Kluding, H. *Wilhelm Roux Arch. dev. Biol.* **193**, 183–295 (1984).
22. Mayer, U. & Nüsslein-Volhard, C. *Genes Dev.* **2**, 1496–1511 (1988).
23. Scholz, H., Deatrick, J., Klaes, A. & Klämbt, C. *Genetics* **135**, 455–468 (1993).
24. Tautz, D. & Pfeifle, C. *Chromasoma* **00**, 81–85 (1989).
25. Gaul, U., Mardon, G. & Rubin, G. M. *Cell* **00**, 1007–1019 (1992).
26. Brand, A. H. & Perrimon, N. *Development* **110**, 401–415 (1993).
27. Basler, K., Yen, D., Tomlinson, A. & Hafen, E. *Genes Dev.* **4**, 728–739 (1990).
28. Basler, K., Christen, B. & Hafen, E. *Cell* **04**, 1069–1082 (1991).
29. Alessi, D. R. *et al. EMBO J.* **13**, 1610–1619 (1994).
30. Traverse, S., Gomez, N., Paterson, H., Marshall, C. & Cohen, P. *Biochem. J.* **200**, 351–355 (1992).

ACKNOWLEDGEMENTS. We thank C. Marshall and S. Cowley for initial help with the phosphorylation assay and for providing the activated MAPKK; G. Rubin for the yan-A cDNA and the anti-Elav antibody; K. Basler, B. Dickson, C. Lehner and J. Campos-Ortega for comments on the manuscript; and C. Hugentobler and A. Klaes for technical assistance. The first three authors contributed equally to this work.

Rapid evolution of a protein *in vitro* by DNA shuffling

Willem P. C. Stemmer

Affymax Research Institute, 4001 Miranda Avenue, Palo Alto, California 94304, USA

DNA SHUFFLING is a method for *in vitro* homologous recombination of pools of selected mutant genes by random fragmentation and polymerase chain reaction (PCR) reassembly[1]. Computer simulations called genetic algorithms[2–4] have demonstrated the importance of iterative homologous recombination for sequence evolution. Oligonucleotide cassette mutagenesis[5–11] and error-prone PCR[12,13] are not combinatorial and thus are limited in searching sequence space[1,14]. We have tested mutagenic DNA shuffling for molecular evolution[14–18] in a β-lactamase model system[9,19]. Three cycles of shuffling and two cycles of backcrossing with wild-type DNA, to eliminate non-essential mutations, were each followed by selection on increasing concentrations of the antibiotic cefotaxime. We report here that selected mutants had a minimum inhibitory concentration of 640 μg ml^{-1}, a 32,000-fold increase and 64-fold greater than any published TEM-1 derived enzyme. Cassette mutagenesis and error-prone PCR resulted in only a 16-fold increase[9].

The poorly hydrolysed antibiotic cefotaxime has a minimum inhibitory concentration (MIC) of only 0.02 μg ml^{-1} for *Escherichia coli* containing the TEM-1 β-lactamase expressed from the vector p182Sfi. The TEM-1 gene was digested into random fragments with DNase I. These small fragments were reassembled into full-length sequences using a PCR-like process and the shuffled sequences reinserted into the vector[1] (Fig. 1). Recombination is caused by the incorporation of a fragment derived from one sequence into another, based on homology. In addition, this method produces a point mutagenesis rate of 0.7%, similar to error-prone PCR[1,13]. This process was repeated for three rounds, and after each round, mutants with improved resistance were selected by plating on increasing levels of cefotaxime. Several hundred colonies from the highest levels of cefotaxime were used as the PCR template for the next round (Fig. 2). Colonies from rounds 1, 2 and 3 were obtained at 0.32–0.64 μg ml^{-1}, 5–10 μg ml^{-1} and 40–80 μg ml^{-1}, respectively. Some colonies from round 3 had a MIC of 320 μg ml^{-1}. Because cefotaxime resistance is cell-density-dependent, the MIC was standardized to 1,000 cells per plate (24 h, 37 °C). At higher cell density, colonies grew at up to 1,280 μg ml^{-1}. A β-lactamase gene (ST-1) of a selected colony contained nine base substitutions, including four silent mutations (Fig. 2).

We attempted to remove all non-essential mutations by backcrossing[1]. ST-1 was shuffled for two rounds in the presence of a 40-fold excess of wild-type DNA fragments (Fig. 2). Small DNA fragments (30–100 bp) were used to increase the efficiency of the backcross. A new β-lactamase gene (ST-2), obtained at 1,280 μg ml^{-1}, had a MIC of 640 μg ml^{-1}. As expected, all four silent mutations had reverted to wild-type sequence. Mutation g4205a, located between the −35 and −10 sites of the β-lactamase P3 promoter, was retained. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS–PAGE) analysis of periplasmic extracts showed that ST-1 and ST-2 express 2–3-

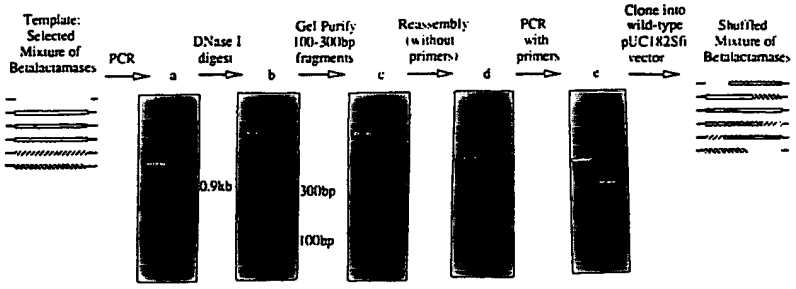
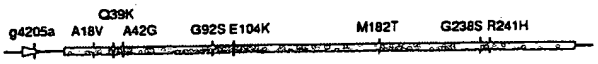


FIG. 1 DNA shuffling of the TEM-1 β-lactamase followed by selection for cefotaxime resistance. A single round of shuffling consists of amplifying the β-lactamase gene by PCR, cutting the PCR product into random fragments with DNase I, gel purification of small fragments, reassembly of the fragments in a PCR-like reaction without primers, amplification of the reassembled product by standard PCR, followed by cloning into the vector and selection on cefotaxime.

METHODS. p182Sfi contains the TEM-1 β-lactamase flanked by *Sfi*I restriction sites. The *Sfi*I sites were added 5' of the promoter and 3' of the end of the gene by PCR of the vector with the primers TTCTATTGACGGCCTGTCAGGCCTCAT-ATATACTTTAGATTGATTT and TTGACGCACTGGCCATGGTGGCCAAAAATAAACAAATAGGGGTTCCGCGCACATTT, and by PCR of β-lactamase gene with two other primers listed below. The substrate for the shuffling reaction was dsDNA of 0.9 kb obtained by PCR of p182Sfi with the primers AACTGACCACGGCCTGACAGGCCGGTCTGACAGTTACCAATGCTT and AACCTGTCCTGGCCACCATGGCCTAAATACATTCAAATATGTAT. In rounds 2 and 3 a mixture of >100 cefotaxime^r colonies was used as the template for the PCR. Colony PCR programme: 10 µl of cells in LB broth, 10 min, 99 °C, 35× (94 °C, 30 s; 52 °C, 30 s; 72 °C, 30 s), 5 min, 72 °C. The removal of free primers from the PCR product by Wizard PCR prep (Promega, Madison, WI) was found to be very important. Afew µg of the DNA substrate was digested with 0.15 units of DNase I (Sigma) in 100 µl 50 mM Tris–HCl pH 7.4, 1 mM $MgCl_2$, for 10 min at room temperature. Fragments of 100–300 bp were purified from a 2% low melting point agarose gel and resuspended in PCR mix (0.2 mM each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris–HCl pH 9.0, 0.1% Triton X-100) at 10–30 ng µl^{-1}. No primers were added at this point. A PCR programme of 94 °C, 1 min, 40× (94 °C, 30 s; 50–55 °C, 30 s; 72 °C, 30 s), was used in an MJ Research PTC-150 (Watertown, MA) thermocycler. After 40-fold dilution of the minus primer product into PCR mix with 0.8 µM of each primer and 15–20 additional cycles of PCR (94 °C, 30 s; 50 °C, 30 s; 72 °C, 45 s), reproducibly a single product of 900 bp is obtained.

TABLE 1 Characterization of cefotaxime resistance of different combinations of mutations

Name	Genotype	MIC	Source of MIC
TEM-1	Wild-type	0.02	This study
—	E104K	0.08	Ref. 9
—	G238S	0.16	Ref. 9
TEM-15	E104K/G238S*	10	This study
TEM-3	E104K/G238S/Q39K	10*	This study
		2–32	Refs 19, 20
ST-4	E104K/G238S/M182T*	10	This study
ST-1	E104K/G238S/M182T/A18V/t3959a/g3713a/g3934a/a3689g*	320	This study
ST-2	E104K/G238S/M182T/A42G/G92S/R241H/t3842c/a3767g*	640	This study
ST-3	E104K/G238S/M182T/A42G/G92S/R241H*	640	This study



The base numbers (small letters) correspond to the revised pBR322 sequence[24], and the amino-acid numbers (capitals) correspond to the ABL standard numbering scheme[25]. Specific combinations of mutations were introduced into the wild-type p182Sfi by PCR, using two oligonucleotides per mutation. The separate PCR fragments were gel purified and combined by overlap PCR using 10 ng of each fragment. PCR was done for 25 cycles without outside primers, followed by 25 cycles in the presence of the *Sfi*-containing outside primers. The oligonucleotides for mutation A42G were AGTTGGGTGGACGAGTGGGTTACATCGAACT and AACCCACTCGTCCACCCAACTGATCTTCAGCAT, for Q39K AGTAAAAGATGCTGAAGATAAGTTGGGTGCACGAGTGGGTT and ACTTATCTTCAGCATCTTTTACTT, for G92S AAGAGCAACTCAGTCGCCGCATACACTATTCT and ATGCGGCGACTGAGTTGCTCTTGCCCGGCGTCAAT, for E104K TATTCTCAGAATGACTTGGTTAAGTACTCACCAGTCACAGAA and TTAACCAAGTCATTCTGAGAAT, for M182T AACGACGAGCGTGACACCACGACGCCTGTAGCAATGGCAA and TCGTGGTGTCACGCTCGTCGTT, for G238 alone TTGCTGATAAATCTGGAGCCAGTGAGCGTGGGTCTCGCGGTA and TGGCTCCAGATTTATCAGCAAT, for G238S and R231H (combined) ATGCTCACTGGCTCCAGATTTATCAGCAAT and TCTGGAGCCAGTGAGCATGGGTCTCGCGGTATCATT, for g4205a AACCTGTCCTGGCCACCATGGCCTAAATACAATCAAATATGTATCCGCTTATGAGACAATAACCCTGATA.

* All these mutants additionally contain the g4205a promoter mutation.

fold more β-lactamase than the wild-type plasmid (data not shown). ST-2 contained three of the four amino-acid mutations of ST-1 (E104K, M182T and G238S), as well as three new amino-acid mutations (c3441t resulting in R241H, c3886t resulting in G92S, and g4035c resulting in A42G) and two new silent mutations (t3842c and a3767g).

For comparison with published data, several combinations of mutations were constructed into the wild-type p182Sfi vector (Table 1). All contained the g4205a promoter mutation, and were confirmed by partial sequencing. The known clinical TEM-1-derivatives (TEM-1–19) all contain up to four of a set of eight dispersed mutations[19–21]. Because the maximum MIC obtained by cassette mutagenesis was only 0.64 µg ml^{-1} (ref. 9), high resistance apparently cannot be obtained by mutagenesis of one area. Mutations E104K or G238S are present in published cefotaxime-resistant TEM-1 derivatives (TEM-3, 4, 6, 8, 9, 14–19), and were obtained separately by cassette mutagenesis, with MICs of only 0.08 and 0.16 µg ml^{-1} (ref. 9). In contrast, a combinatorial mutagenesis approach might have yielded the double mutant (TEM-15; ref. 19) with a MIC of 10 µg ml^{-1}. These two mutations thus appear to be synergistic. E104K and G238S in combination with Q39K (TEM-3) or T263M (TEM-4) have reported MICs of 2–32 µg ml^{-1} (refs 19, 21), whereas a TEM-3 like construct in our vector had a MIC of 10 µg ml^{-1}. A construct containing the three amino-acid changes that were conserved after the backcross (E104K, M182T, G238S) also had a MIC of 10 µg ml^{-1}. With or without the silent mutations, constructs containing all of the six amino-acid changes of ST-2 introduced into the wild-type gene yielded colonies with the same MIC as ST-2 (640 µg ml^{-1}). Thus, the six amino-acid mutations (plus the promoter mutation) conferred the high-resistance phenotype.

A control experiment using error-prone PCR but no shuffling resulted in a MIC of only 0.32 µg ml^{-1} after three selection cycles.

A useful approach may be to shuffle many related, naturally occurring genes, such as antibodies[22,23] or homologous genes from different species. The diversity present in such a mixture may be more meaningful than random mutations. □

Received 16 February; accepted 21 June 1994.

1. Stemmer, W. P. C. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
2. Holland, J. H. *Scient. Am.* **267**, 66–72 (1992).
3. Holland, J. H. *Adaptation in Natural and Artificial Systems* 2nd edn (MIT Press, Cambridge, 1992).

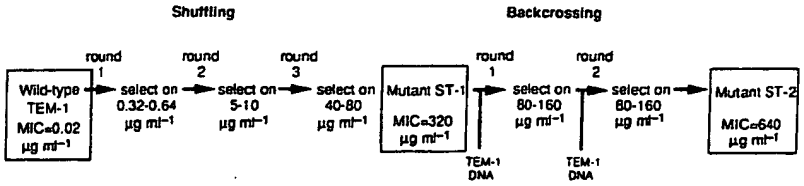


FIG. 2 Three successive rounds of DNA shuffling were done and the cells were grown on increasing cefotaxime levels. The MIC of cefotaxime (Sigma) for *E. coli* XL1-blue (Stratagene, San Diego) carrying wild-type p182Sfi is 0.02 μg ml^{-1}. A mutant with a 16,000-fold increased resistance to cefotaxime was obtained (MIC = 320 μg ml^{-1}). This mutant was backcrossed twice, by shuffling with a 40-fold excess of wild-type DNA. The backcrossed mutant was 32,000-fold more resistant than the wild type (MIC = 640 μg ml^{-1}). After selection, the plasmid of selected clones was transferred back into wild-type XL-1 blue cells to ensure that none of the measured drug resistance was due to chromosomal mutations. DNA sequencing showed that both mutants had 9 single-base-pair mutations.

7. Arkin, A. & Youvan, D. C. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7811–7815 (1992).
8. Delagrave, S. & Youvan, D. C. *Biotechnology* **11**, 1548–1552 (1993).
9. Patzkill, T. & Botstein, D. *J. Bact.* **174**, 5237–5243 (1992).
10. Oliphant, A. R., Nussbaum, A. L. & Struhl, K. *Gene* **44**, 177–183 (1986).
11. Hermes, J. D., Blacklow, S. C. & Knowles, J. R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 696–700 (1990).
12. Leung, D. W., Chen, E. & Goeddel, D. V. *Technique* **1**, 11–15 (1989).
13. Caldwell, R. C. & Joyce, G. F. *PCR Meth. Applic.* **2**, 28–33 (1992).
14. Kauffman, S. A. *The Origins of Order* (Oxford Univ. Press, New York, 1993).
15. Kauffman, S. A. *J. theor. Biol.* **157**, 1–7 (1992).
16. Bartel, D. P. & Szostak, J. W. *Science* **261**, 1411–1418 (1993).
17. Tuerk, C. & Gold, L. *Science* **249**, 505–510 (1990).
18. Joyce, G. F. *Scient. Am.* **267**, 90–97 (1992).
19. Jacoby, G. A. & Medeiros, A. A. *Antimicrob. Ag. Chemother.* **35**, 1697–1704 (1991).
20. Collatz, E., Labia, R. & Gutmann, L. *Molec. Microbiol.* **4**, 1615–1620 (1990).
21. Philippon, A., Labia, R. & Jacoby, G. *Antimicrob. Ag. Chemother.* **33**, 1131–1136 (1989).
22. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. *Nature* **348**, 552–554 (1990).
23. Huse, W. D., Sastry, L., Iverson, S. A. & Kang, A. S. *Science* **246**, 1275–1278 (1989).
24. Watson, N. *Gene* **70**, 399–403 (1988).
25. Ambler, R. P. *et al. Biochem. J.* **276**, 269–272 (1991).

ACKNOWLEDGEMENTS. We thank A. Crameri for assistance with DNA sequencing, P. Schatz for suggesting the model system and W. Dower for valuable discussions.

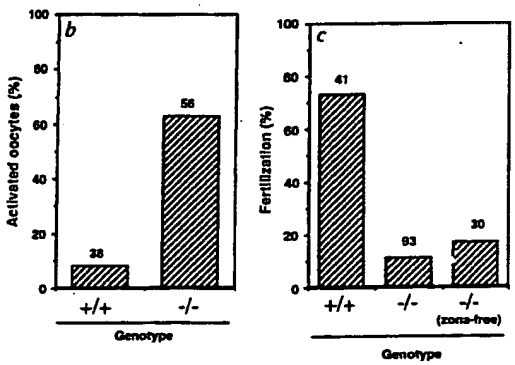
ERRATA

Parthenogenetic activation of oocytes in c-*mos*-deficient mice

Naohiro Hashimoto, Nobumoto Watanabe, Yasuhide Furuta, Hiroyuki Tamemoto, Noriyuki Sagata, Minesuke Yokoyama, Kenji Okazaki, Mariko Nagayoshi, Naoki Takeda, Yoji Ikawa & Shinichi Aizawa

Nature **370**, 68–71 (1994)

FIGURE 3*b* and *c* of this Letter was an early version that should not have been published. The correct version of this figure is shown here. □



Degradation of trifluoroacetate in oxic and anoxic sediments

Pieter T. Visscher, Charles W. Culbertson & Ronald S. Oremland

Nature **369**, 729–731 (1994)

IN the last sentence of the opening paragraph of this Letter, an error was introduced during editing in which fluoroform was referred to as a "potential ozone-depleting compound." In fact, fluoroform as well as other HFCs were recently shown by Ravishankara *et al.*[1] to have "negligibly small" ozone depletion potentials. □

1. Ravishankara, A. R. *et al. Science* **263**, 71–75 (1994).

Miller-Dieker lissencephaly gene encodes a subunit of brain platelet-activating factor acetylhydrolase

Mitsuharu Hattori, Hideki Adachi, Masafumi Tsujimoto, Hiroyuki Arai & Keizo Inoue

Nature **370**, 216–218 (1994)

THE word 'acetylhydrolase' was accidentally omitted from the end of the title of this paper. The correct title should read "Miller-Dieker lissencephaly gene encodes a subunit of brain platelet-activating factor acetylhydrolase". □

Attachment D

The Human Immunoglobulin Genes

T.H. RABBITTS

Laboratory of Molecular Biology, The Medical Research Council Centre, Hills Road, Cambridge CB2 2QH, U.K.

The Nineteenth Colworth Medal Lecture

Delivered on 23 September 1982 at the University of Aberdeen

Dr. T. H. RABBITTS

Introduction

The results that I will describe here concern the structure and rearrangement of the genes coding for antibodies in man. A great deal of work was conducted in several laboratories on the analogous genes of mouse. I shall make no attempt to review this work. The human antibody gene system was chosen for study largely because of the existence of a wide range of defined, clinical disorders of the immune system, which include the immunodeficiency diseases (e.g. agammaglobulinaemia) and the leukaemias [frequently tumours of immunoglobulin (Ig) producing cells]. In the preparation and analysis of molecular probes for the immunoglobulin genes, we have gained some knowledge of the arrangement and rearrangement of these genes in man and it is clear that the human antibody genes represent a fluid system which employs a number of highly developed DNA rearrangement procedures in their activation.

The antibody proteins are made up from two types of polypeptide chain, the heavy (H) and light (L) chains. Each polypeptide itself has an *N*-terminal variable (V) region and a *C*-terminal constant (C) region. The V-region of the heavy (V_H) and light (V_L) chains together make the antibody combining site capable of specific antigen recognition (defined by the amino-acid sequence of V-regions). The C-region, particularly that of the H-chain (C_H), performs more constant functions such as complement fixation. It is the C_H-region which defines the antibody class; there are five classes of C_H-sequence called μ, δ, γ, ε or α (giving IgM, IgD, IgG, IgE and IgA respectively). The C_H-region sequence is invariant in each class except γ and α where amino-acid differences define γ_1, γ_2, γ_3 or γ_4 and α_1 or α_2 (giving IgG1, 2, 3 or 4 and IgA1 or 2 respectively). Any of the H-chain classes or subclasses can associate with either of the two types of L-chain, κ- or λ-chains (again defined by their respective amino-acid sequences).

During the development of a lymphocyte the first gene to be expressed is μ followed by L-chain induction resulting in the formation of surface IgM (Cooper *et al.* 1976; Knapp *et al.* 1973; Pernis *et al.*, 1976). A B-lymphocyte clone can subsequently initiate IgD production, in addition to the IgM it already makes; the two heavy chains involved here (i.e. μ and δ) express the same V_H-segment A subsequent event occurs, the H-chain class switch, which results in the expression of IgG, IgA or IgE in place of IgM and IgD but maintaining the same antibody combining site (Sledge *et al.* 1976; Wang *et al.* 1970). These various events involve a complex set of chromosomal DNA rearrangements. Basically in the germ-line the antibody genes are in pieces and the fully active gene is created, within the B-cells, by these rearrangements.

Chromosomal mapping of the human Ig genes and association to sites of translocation in malignant lymphoid cells

The chromosomal assignments of the H- and L-chain genes have been made by a variety of methods and the results of our studies and those of other laboratories are summarized in Table 1. We showed that V_H- and C_H-genes both reside in chromosome 14 by using Southern filter hybridization techniques (Southern, 1975) to analyse the presence of H-chain genes in mouse × human somatic cell hybrids which have a defined but incomplete human karyotype (Hobart *et al.* 1981). Fig. 1 is a diagram of this analysis in which positively hybridizing clones were scored and the data clearly show that the H-chain genes occur on chromosome 14.

In accord with the work of Erikson *et al.* (1981) showing that chromosome 22 carries the λ light chain locus, a clone library had been prepared from chromosome 22 which had been

Table 1. *Chromosomal mapping of human immunoglobulin genes*

Gene locus	Assigned chromosome	Reference
κ-chain	2	Malcolm *et al.*, 1982 McBride *et al.*, 1982
λ-chain	22	Erikson *et al.*, 1981 McBride *et al.*, 1982 Rabbits *et al.*, 1982
H-chain	14	Croce *et al.*, 1979 Hobart *et al.*, 1981 Kirsch *et al.*, 1982

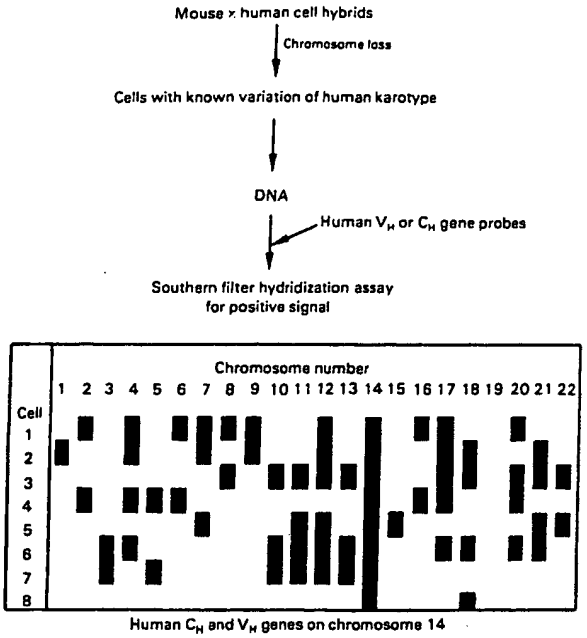


Fig. 1. *Chromosomal localization of human heavy chain genes*

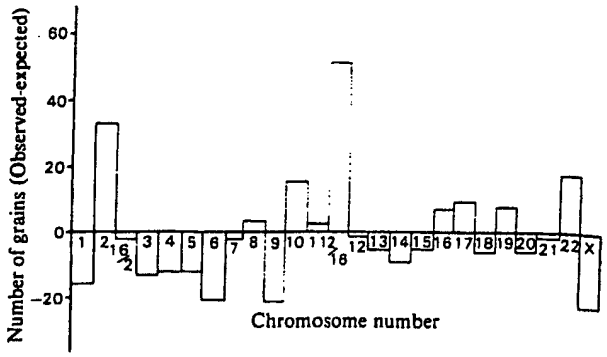


Fig. 2. *In situ hybridization of [^{3}H]cRNA, made on the HK101 gene, to translocated (2;16) chromosomes*

The number of grains observed over all chromosomes from 22 karyotyped cells hybridized to HK101 is shown. The expected value for grains is calculated by dividing total grains counted by the relative chromosome size.

fractionated in the fluorescence activated cell sorter (Krumlauf *et al.*, 1982) and this library was used to isolate both V_λ and C_λ genes (T. H. Rabbitts, F. T. Kao & B. Young, unpublished work); the proof of concordance between these λ-genes and the long arm of chromosome 22 was made by showing a selective loss of the relevant genes in cell-hybrids with progressive deletions of the long arm of chromosome 22 (T. H. Rabbitts, F. T. Kao & B. Young, unpublished work).

A completely different approach was used to localize the κ-chain genes to chromosome 2. This was the method of *in situ* hybridization using cloned V_κ probes (Malcolm *et al.* 1982). The hybridization of two independent genomic V_κ clones to metaphase spreads of phytohaemagglutinin-stimulated T-lymphocytes or to fibroblast cultures showed a specific localization of grains to chromosome 2. The specificity of such hybridization can be shown in a variety of ways but one in which the gene localization is more clearly established is to study the hybridization to cells carrying balanced reciprocal translocations. Fig. 2 shows the results of V_κ-probe hybridization to a cell carrying a 2;16 translocation [46XXt(2;16)(q13;q22)]. The specific hybridization signal was observed on the normal chromosome 2 and in the 2/16 chromosome, whilst no signal could be seen in association with either 16/2 or the normal 16 chromosome. The hybridization to chromosome 2/16 further helps to localize the κ-genes within either the short arm or the section of the long arm between q13 and the centromere of chromosome 2. In fact an analysis of the grain distribution over all the normal chromosomes 2 in 106 karyotypes enabled us to localize the V_κ hybridization to the short arm near the centromere (2p12).

The localization of the three immunoglobulin genes to the three separate autosomes 2 (κ), 14 (H) and 22 (λ) has some significant implications for some specific chromosome translocations identified in the malignant lymphoid tumours, Burkitt lymphoma and chronic myeloid leukaemia. Many Burkitt lymphoma cells show a specific 8;14 translocation (Klein, 1981) whilst other variant cells show either 2:8 or 8;22 translocations (Bernheim *et al.*, 1981). Furthermore, greater than 90% of chronic myeloid leukaemia patients possess Ph' positive cells which seem to be reciprocal translocation of chromosomes 22 and 9 (Rowley, 1982). It is interesting that in Burkitt lymphoma each of the Ig gene-containing chromosomes can be involved in a specific translocation of apparently the same segment of chromosome 8 and also that the break in the chromosome 22 of chronic myeloid leukaemia is related to the analogous Burkitt's lymphoma break.

The specificity of the chromosomal translocations in these malignant tumours may well be coupled to the fact that the break points are associated in some way with the Ig genes which are known to be undergoing rearrangements during their activation. This raises two possibilities. Firstly that the translocation in itself is an event which contributes to the cell transformation and secondly that Ig gene rearrangement together with related sequences are responsible for the occasional inter-chromosomal event which has such catastrophic consequences. In the remainder of this paper, the different types of DNA rearrangements and the associated sequences which occur in the human Ig locus will be discussed.

The structure and origin of diversity of human V-genes

The germ-line DNA contains separate V- and C-region genes which join in the expressing cells (Hozumi & Tonegawa, 1976; Rabbitts & Forster, 1978). Isolated V_κ- and V_H-genes have a structure outlined in Fig. 3. Both types of V-gene have a 5′ leader sequence interrupted at the codon −5 by an intervening sequence of about 100 nucleotides (Matthyssens & Rabbitts, 1980; Bentley & Rabbitts, 1980). The remainder of the sequence is uninterrupted coding region. Both genes end prematurely in the germ-line compared with the V-region as defined by the protein sequence. This excluded portion of the V-region is the J-region (and also D-region in H-chains) which lies adjacent to the C-gene (see the next section).

The portion of the antibody which recognises antigen is the V-region, so that a multiplicity of different V-gene sequences are required. How are these sequences generated? One component of the diversity of V-genes is undoubtedly the presence of multiple V-genes in the germ-line. This can simply be demonstrated by Southern filter hybridization analysis of nuclear DNA using V-gene probes. An example of this is shown in Fig. 4 in which the hybridization probe was derived from the V_κI subgroup. Each of eight unrelated DNAs displayed about 15 hybridization bands with evidence of some genetic polymorphism. V_H probes used in analogous experiments showed similar patterns of bands although the complexity of the hybridization observed was greater (Matthyssens & Rabbitts, 1980). When the hybridization properties of probes from two

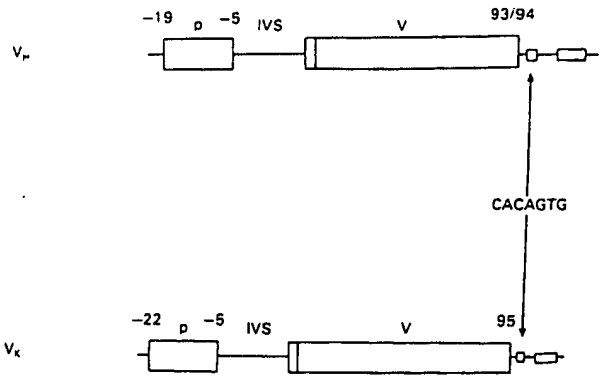


Fig. 3. *Structure of human V-genes*

V_κ and V_H-genes were isolated from foetal liver DNA. p, Precursor or leader sequence; IVS, intervening sequence; V, V-segment or V-gene.

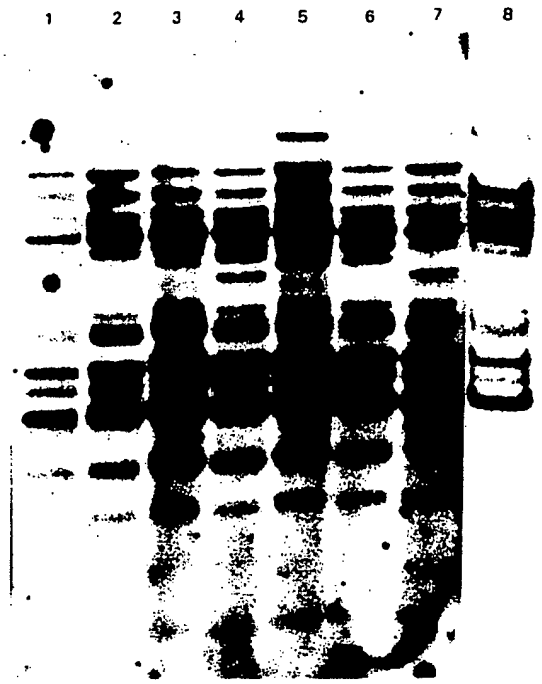


Fig. 4. *Southern filter hybridization patterns of V_κ-genes from eight unrelated humans*

A V_κI probe was hybridized with *BglII*-digested DNA from: slot 1, 160 cell-line (Karpas *et al.*, 1977); slot 2, MOLT4 cell-line (Minowada *et al.*, 1972); slot 3, DAUDI cell-line (Klein *et al.*, 1968); slot, 4. HeLa cells; slot 5, spleen; slot 6; placenta; slot 7, placenta; slot 8, foetal liver.

different human V_κ subgroups were compared, we found very little difference in the patterns of hybridization (Bentley & Rabbitts, 1981) indicating that we were detecting, with a given probe, a large proportion of the V_κ-gene pool. This conclusion has implications for the origin of V-gene diversity in man. The size of the V_κ-gene pool based on these estimates is likely to be of the order of 20–30 different V-genes per chromosome. If we assume that allelic variation is widespread we can double these figures (i.e. 40–60). This limited number of V_κ-genes indicates that somatic variation is a major contributor to the total V_κ repertoire in man. The arguments supporting such a conclusion have been discussed (Bentley & Rabbitts, 1981) so will not be repeated here.

Joining segments of the human H-chain locus

In addition to the mutational events alluded to above there is a combinatorial diversity [first described for mouse κ-chains by the laboratories of Leder and Tonegawa (Max *et al.*, 1979; Sakano *et al.*, 1979)] generated by the V-gene joining which occurs during the active gene formation. As mentioned above, both V_H- and V_κ-genes in the germ-line do not encode the full V-region as defined by comparative protein sequencing of many V-regions. This deficit is made up in part from a set of J_H or joining segments located upstream of the C_μ-gene. We identified four J_H and a pseudo (ψ) J_H segment about 8000 bases upstream of the 5′ end of the human μ-gene. Two further J_H-regions have been identified (Ravetch *et al.*, 1981). The sequence and arrangement of these is shown in Fig. 5 where the region encoding J_H, C_μ and C_δ is indicated as deduced from two overlapping phage clones derived from human foetal liver DNA. The DNA which comes between the various J_H segments (about 300 bases) and between the J_H- and μ-gene is intervening sequence which is post-transcriptionally removed. Similarly the μ-gene, in keeping with all C_H-genes in mouse and man, has genetic domains (reflecting the protein domains) separated by short intervening sequences (around 300 bases in length). The human μ-gene, like the ε-gene (Rabbitts *et al.*, 1981; Flanagan & Rabbitts, 1982*a*), but unlike the γ-genes (Krawinkel & Rabbitts, 1982; Ellison & Hood, 1982; Takahashi *et al.*, 1982), possesses four distinct domains and lacks a hinge segment. In the human γ-genes, each has a single hinge segment between $C_\gamma 1$ and $C_\gamma 2$ domains; the exception is the γ_3-gene which has four separate hinge segments lying between $C_\gamma 1$ and $C_\gamma 2$ (Krawinkel & Rabbitts, 1982; Takahashi *et al.*, 1982).

Whereas the V_H-gene segment has codons 1–93 the J_H-segment only carries the codons 100–113. This means that even including these segments a small piece of V-region is missing: this segment is the D-segment (first postulated in the mouse by Early *et al.*, 1980*a*). Such D-segments have subsequently been identified in human DNA (Siebenlist *et al.*, 1981) making a complete picture of the V_H-gene integration possible; this is diagrammatically shown in Fig. 6. One of a set of V_H-segments joins to one [or possibly more (Kurosawa & Tonegawa, 1982)] D-segments followed by joining to a J_H-segment. These multiple chromosomal rearrangements lead to the formation of the active C_μ-gene. The cells are now able to make μ-chains which, of course, come from a transcript containing V_H and C_μ on a single pre-mRNA molecule. RNA splicing subsequently removes transcribed intervening sequences (Rabbitts, 1978) resulting in the mRNA from which heavy chain protein is translated.

The events described above indicate that a remarkable set of DNA rearrangement events take place which lead to the transcription of the active μ-gene in association with the integrated V-gene. One of the puzzles of this gene activation is that in the B-lymphocyte the germ-line V-genes are not transcribed (Mather & Perry, 1981). This indicates a very powerful mechanism which selects for the integrated gene. Such a mechanism could be simply at the level of DNA sequence; for example, new sequences could be created or old ones removed after V-gene joining, thereby creating an active promoter. An alternative possibility is that all germ-line V-genes carry a promoter sequence (which is transported with the gene during integration) but that the non-integrated genes are masked from transcription (e.g. by the chromatin structure). In relation to these possibilities, we asked whether or not germ-line V_κ-genes carry a promoter sequence and the answer seems to be that they do (Bentley *et al.*, 1982). We have utilized cloned V_κ-genes from human foetal liver (i.e. unintegrated, non-expressed V-genes) to study their ability to undergo transcription in either an *in vitro* system prepared from HeLa cells or *in vivo* after injection into

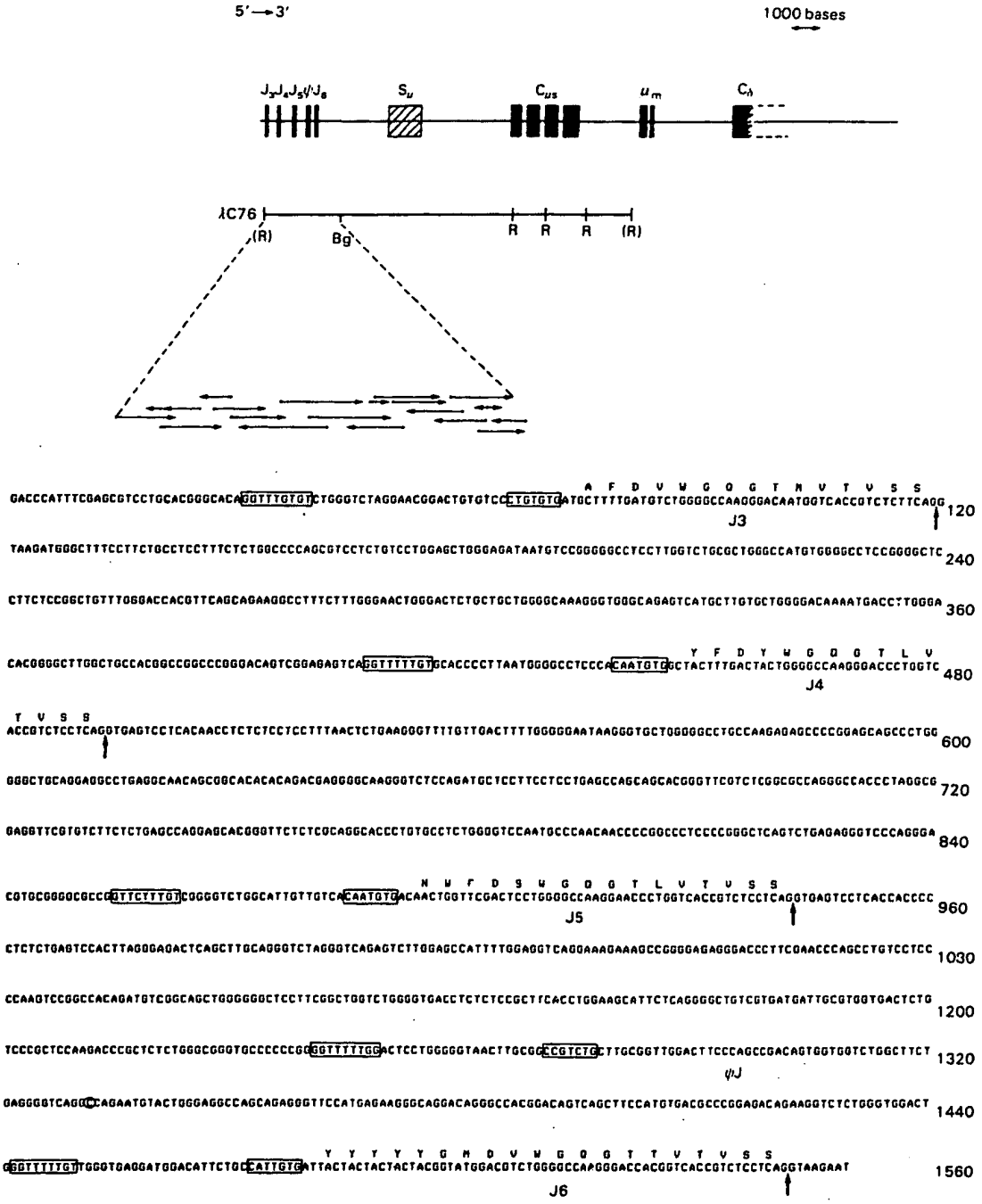


Fig. 5. *Structural arrangement of the human genome encoding J_H-, μ- and δ-genes*

Two recombinant phage clones carrying human foetal liver DNA were isolated by their hybridization to a mouse μ probe. Restriction mapping and nucleotide sequencing revealed the genomic structure shown in the top line (S = switch segment). The sequence of the coding regions of the J_H-segments are indicated and the RNA splicing sites are arrowed. The boxed sequences are signals thought to be important for the process of V–D–J joining.

Xenopus oocytes. We found that two randomly selected V_κ-genes were perfectly adequate templates for specific RNA synthesis in these systems. The transcription was found to start at a point approximately 30 bases downstream of a 'TATA' box (Corden *et al.*, 1980) and 20 bases upstream of the protein initiation codon. A third V_κ-pseudo-gene (which apparently lacks appropriate signal sequence; see later) was unable to support specific RNA synthesis. We conclude from these experiments that sequences necessary for the transcription of mRNA are present on germ-line V_κ-genes. This shows that DNA rearrangement as such is not a pre-requisite for V-genes to be capable of transcription. Taken together with the observation

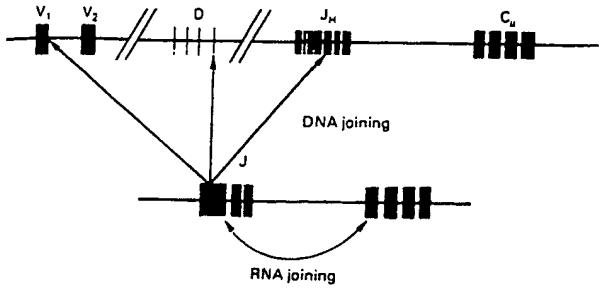


Fig. 6. *Proposed events in the formation and transcription of the active heavy chain μ-gene*

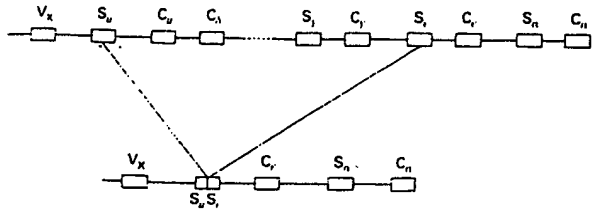


Fig. 7. *Scheme for the process of switch recombination*

that long stretches upstream of integrated mouse V-genes appear identical with germ-line counterparts argues that other factors than the immediate 5′ sequence controls V-gene transcription.

The H-chain class switch: deletion of C_H-genes

After integration of the V_H/D/J segments, μ-chains appear in the cytoplasm of cells without L-chain. Subsequently L-chain integration occurs and the cells express surface IgM and can also express IgD. These cells carry the same V_H segment associated with the δ- or μ-chains. A further differentiation pathway, the H-chain class switch, takes a V_H-gene away from the C_μ-gene and this V_H is transcribed with C_γ, C_α or C_ε. Studies on the arrangement of C_H-genes in mouse myeloma cells expressing various classes of Ig showed that in plasma cells the switch is accompanied by deletion of genetic material between J_H and the newly expressed C_H-gene (Honjo & Kataoka, 1978; Cory & Adams, 1980; Coleclough *et al.*, 1980; Rabbitts *et al.*, 1980*a*). This deletion includes C_H-genes in the interim region. Detailed analysis of cloned C_H-genes from mouse myelomas revealed that this deletion was taking place between switch or S-segments located just upstream of each C_H except C_δ. The process of such a deletion is summarized in Fig. 7, which draws from our own work on mouse and human C_H-genes and that of various other laboratories (Davis *et al.*, 1980: Kataoka *et al.*, 1980; Sakano *et al.*, 1980). The switch recombination process seems to involve sequences within the S-segments which are located near each C_H-gene and which are homologous to each. Our studies of a mouse myeloma mutant IgG1 chain (Dunnick *et al.*, 1980) first identified short tandemly repeated regions apparently present in each C_H S-region which undergo rearrangements in the class switch. These segments were proposed to be mediators of S–S recombination, in a process such as unequal crossing over between sister chromatids (Fig. 8). This unequal crossing over would result in deletion of material and a switched genotype on one resulting chromosome (Rabbitts *et al.*, 1980*b*).

The involvement of these S-segments in recombination, of course, requires a sufficient homology between those of each C_H-gene. Such homology can be shown by hybridization and indeed such hybridization profiles reveal some interesting features of the relative homologies of different S-segments. Fig. 9 shows the hybridization of a human S_μ probe with four γ-genes (putatively the active genes), the active ε-gene plus a ψε-gene ($\psi\varepsilon_1$, which will be discussed later) and two α-genes (probably α_1 and α_2). Clearly all the C_H-genes analysed display some homology to the S_μ probe. By far the strongest homology is found in the α-genes and the active ε-gene. The $\psi\varepsilon_1$-gene does possess an S-sequence with significant homology to S_μ; in fact, this homology is better than that of S_μ with the most homologous S_γ (i.e. $S_\gamma 3$). In general the S_γ hybridization was weak but none the less significant (Flanagan & Rabbitts, 1982*b*). This result is consistent with the similar result found in mouse and lends support to the theory that S_γ–S_γ recombination is by far the most likely after S_μ–S_γ switching (Marcu *et al.*,1982).

Cotranscription of μ- and δ-genes as a means of C_H switching

We can conclude that switching of C_H-genes in plasma cells generally occurs by recombination between homologous S-segments and resulting in gene deletion. When we analysed the human δ-gene for an S-segment we found no evidence for such a sequence (Rabbitts *et al.*, 1981) so that an alternative mechanism was clearly operating in this case. Mapping of the μ- and δ-genes showed that the δ-gene was located about 5000 bases downstream of the μ-gene and in the same direction of transcription (Fig. 6). Roughly in the middle of the DNA separating these genes we located the coding regions for the membrane segment of μ (Rabbitts *et al.*, 1981) which are exactly analogous with those first described in mouse (Early *et al.*, 1980*b*). The proposal has been made that the membrane and secreted forms of μ result from differential RNA splicing routes which generate one or other type of mRNA (Early *et al.*, 1980*b*). The proximity of the δ- and μ-genes together with the absence of a detectable S-sequence argues that the μ- and δ-genes can be included in a single long transcript from which μ or δ mRNA is made containing the same V_H sequence. The inherent problem with such a scheme derives from the constrains on the RNA splicing mechanism, since in order to produce a VDJ δ mRNA, a set of RNA splicing sites has to be ignored (Milstein *et al.*, 1981).

Pseudo-genes in the human Ig locus

We have studied three ψ-Ig genes which include a ψV_κ-gene (Bentley & Rabbitts, 1980), a ψγ-gene (Krawinkel & Rabbitts, 1982), a ψε-gene (Flanagan & Rabbitts, 1982*b*) and a ψJ_H-segment (Flanagan & Rabbitts, 1982*a*). The structure of each ψ-gene precludes the possibility that they could function as active immunoglobulin genes and a number of interesting, different defects are found in the various genes. Each of the genes has a defective RNA splice site or at least a sequence at variance with the consensus splicing signal. Donor sites–are affected in ψV_κ, J_H [two further ψJ_H with similar properties have also been identified (Ravetch *et al.*, 1981)] and $\psi\varepsilon_1$, whilst an acceptor site is affected in ψγ. The ψJ_H segments do not correspond to any known protein and several insertions/deletions place protein translation termination signals in phase. Similar insertion and deletion mutations were observed in the ψV_κ-gene which again potentially place in phase protein termination signals, even though we could identify the gene sequence as being closely related to the $V_\kappa I$ subgroup. The ψV_κ was also interestingly defective in signals for RNA transcription [i.e. no good match to the consensus TATA box sequence was evident and, indeed, this gene was not transcribable *in vitro* (Bentley *et al.*, 1982)] and for V–J joining. Thus this ψ-gene is probably incapable of joining to a J-segment but, even if it were, transcription seems unlikely. In any event, transcription of this sequence would yield a pre-mRNA which would not be processed nor translatable into a full length protein. The most likely view of such ψ-genes is that they are relics of evolutionary drift which have gone beyond the limits of the constraint which may be imposed by function.

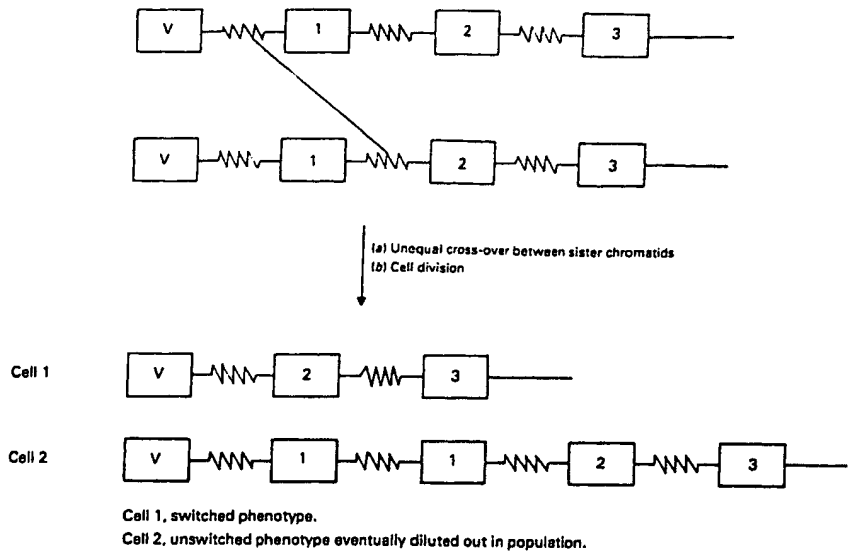


Fig. 8. *Model for switch recombination by sister chromatid exchange*

Genes 1 to 3 represent sequential C_H-genes. S-segments are represented as wavy lines, This is taken from Rabbitts *et al.* (1980*b*)

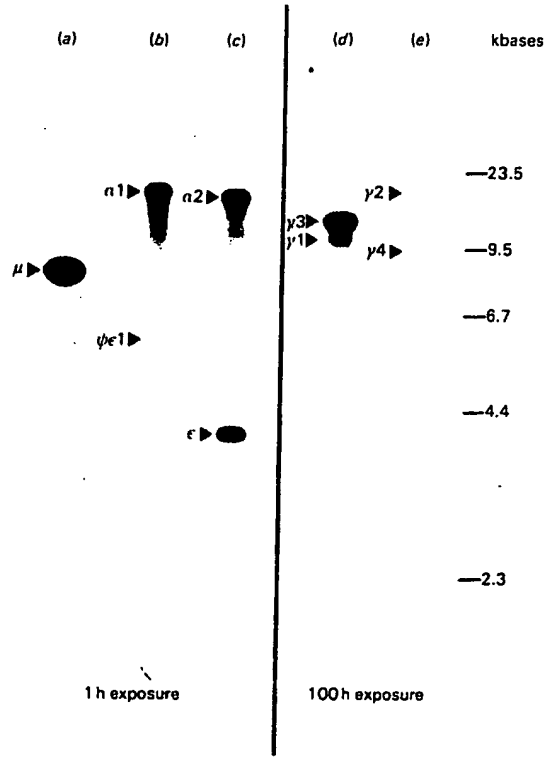


Fig. 9. *Southern filter hybridization between S_μ- and S-segments from other C_H-genes*

An S_μ probe was hybridized to restriction digests of cosmid clones carrying the various C_H-genes indicated in the Figure. After hybridization the filter was autoradiographed for the times indicated.

The $\psi\gamma$-gene, again, does not correspond to any known protein and apart from the known abnormality of a splice signal, may not possess an S-sequence (Takahashi *et al.*, 1982) thereby making it likely that it cannot normally be expressed. Finally the $\psi\varepsilon_1$-gene is unusual in that it has lost the Cε1 and Cε2 domains completely and the first four residues of the Cε3 domain, which therefore includes the RNA splicing signal. Just in front of the site of deletion in this gene a sequence occurs which has a strong relationship with S-sequences. This type of deleted ψ-gene is reminiscent of the *in vitro* isolated IF2 mutant gene which we previously described (Dunnick *et al.*, 1980). The IF2 cell-line was discovered as an isoelectric focusing mutant, derived from the IgG1-producing mouse myeloma X63, which was subsequently shown to lack the complete C_H1 domain (Secher *et al.*, 1973). An analysis of the gene from IF2 showed that an extensive gene deletion had removed the C_H1 coding segment along with most of the C_H1 to hinge intervening sequence and a large part of the V_H to C_H1 intervening sequence (Dunnick *et al.*, 1980). The mutant gene therefore has a structure which facilitated RNA splicing from the donor site of the V-region to the acceptor site of the hinge. A secretable H-chain was thus produced. The sequence of the gene showed that deletion had brought S-sequences very close to the hinge coding segment and thus the structures of $\psi\varepsilon_1$ and IF2 are very similar. The proximity of S-like sequences to the sites of deletion in these two genes is intriguing and suggests the involvement of these sequences in the formation of these mutant sequences. If this is true S-sequence must be important both in gene evolution and in derivations of mutant sequences in B-cells.

Arrangement of human heavy chain genes

The organization of the heavy chain locus in man is important as this region is expressed in a variety of malignant leukaemias. We wanted to be able to compare this region in normal and abnormal DNAs so a map of this gene region was required. We have used both V_H and C_H-region probes to analyse the heavy chain locus. Analysis of various human DNAs with a V_HIII probe indicated a complex picture of around 20 detect-

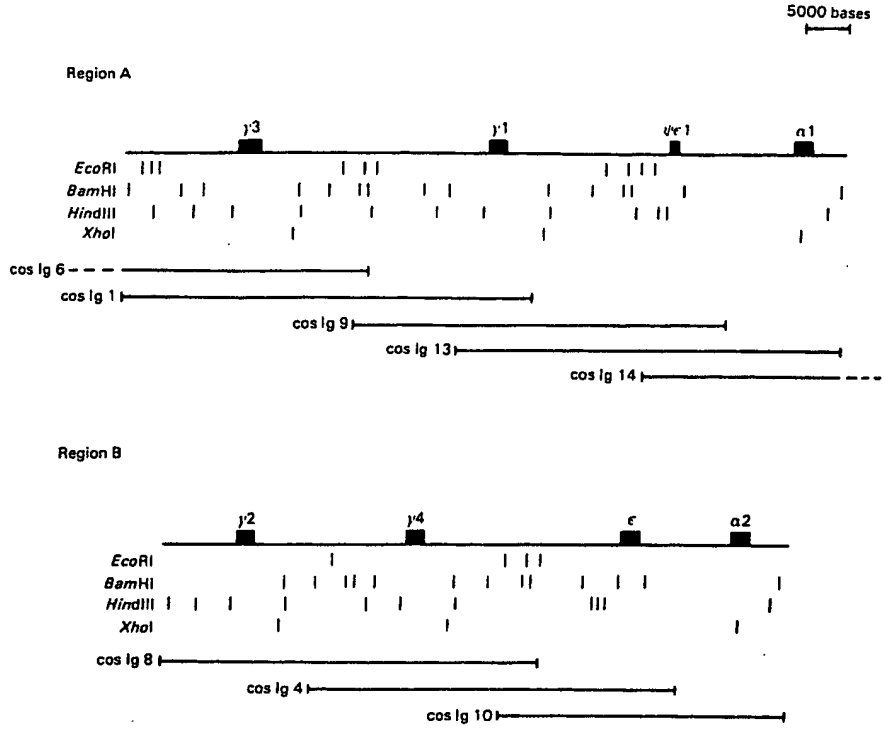


Fig. 10. *Partial map of the human C_H-gene region*

Cosmid clones were isolated by hybridization with γ and ε probes and restriction maps prepared. The broken lines at the ends of groups refer to the regions in which only a single clone has been mapped.

able bands (Matthyssens & Rabbitts, 1980) and the analysis of this region is in its early stages. We studied the arrangement of V_H-segments in five randomly selected genomic clones and determined that the V-genes are separated from one another by quite a large distance; one clone contained two V_H-genes at a distance of 12500 bases whilst a second clone contained two V_H-genes about 15000 bases apart. The other clones analysed contained only one V-gene with the maximum size of genomic DNA, lacking a detectable V-gene, being about 16000 bases. Given an average spacing of 12000 bases and the existence of 100 V_H-genes, this implies that the V_H locus of chromosome 14 occupies at least 1200000 bases.

Recent studies of the human C_H-genes indicate that these genes also encompass a large region of chromosome 14. We have analysed C_H-genes by isolating cosmid clones from a human placental DNA library (prepared by F. Grosveld) and the structure of these clones revealed an interesting organization of genes. Fig. 10 shows the restriction maps of two non-overlapping groups of clones (each containing about 80000 bases of the genome) in which we have placed the relative positions of γ-, ε- and α-genes. The outstanding feature is that both groups of genes possess an order γ–γ–ε–α, implying that the evolution of this locus involved a vast duplication, probably including a segment with the arrangement γ–γ–ε–α or γ–ε–α. This latter duplication was presumably followed by further duplication of γ-genes at each site. This interpretation is more compelling than the other possibility that the initial large duplication included γ–γ–ε–α, since comparison of restriction maps and sequences of the pairs of adjacent γ-genes indicates their close relationship. A complication to this general scheme is the existence of the $\psi\gamma$-gene and to a lesser extent by the unusual quadruplicated hinge segments found in the γ_3-gene (Krawinkel & Rabbitts, 1982; Takahashi *et al.*, 1982). Interestingly the first hinge of the γ_3-gene is closely related to the $\psi\gamma$ hinge, implying an evolutionary relationship between γ_3 and $\psi\gamma_1$. The evolutionary origin of $\psi\gamma_1$ will not become clear until it is placed relative to the other γ-genes and even then the peculiarity of its sequence may preclude a clear conclusion regarding its origin. The study of the C_H-genes in other primates will be very interesting from the point of view of the nature of gene duplications and the origin of the ψ-genes.

Conclusions

The immunoglobulin locus of man is a highly fluid gene system which employs a variety of gene alteration events, in the maturation of the B-lymphocyte, resulting in expansion of diversity in the system. This system has evolved a series of signals which allow chromosomal rearrangement to occur in a specific way. However, it now seems that the very systems which contribute so elegantly to antibody activity may also result in the very occasional specific chromosomal translocations resulting in malignant leukaemias such as Burkitt lymphoma.

I acknowledge the efforts of my various collaborators in the work described here and I am extremely grateful to Alan Forster for his technical expertise. I also thank Dr. C. Milstein and Dr. F. Sanger for help and encouragement.

Bentley, D. L., Farrell, P. J. & Rabbitts, T. H. (1982) *Nucleic Acids Res.* **10**, 1841–1856

Bentley, D. L. & Rabbitts, T. H. (1980) *Nature* (*London*) **288**, 730–733

Bentley, D. L. & Rabbitts, T. H. (1981) *Cell* **24**, 613–633
Bernheim, A., Berger, R. & Lenoir, G. (1981) *Cancer Genet. Cytogenet.* **3**, 307–315
Coleclough, C., Cooper, D. & Perry, R. P. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1422–1426
Cooper, M. D., Kearney, J. F., Lydyard, P. M., Grossi, C. E. & Lawson, A. R. (1976) *Cold Spring Harbor Symp. Quant. Biol.* **41**, 139–191
Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) *Science* **209**, 1406–1414
Cory, S. & Adams, J. M. (1980) *Cell* **19**, 37–51
Croce, C. M., Shander, M., Martinis, J., Cicurel, L., D'Amona, G. G., Dolby, T. W. & Koprowski, H. (1979) *Proc. Natl. Acad. Sic. U.S.A.* **76**, 3416–3419
Davis, M. M., Calame, K., Early, P. W., Livant, D. L., Joho, R., Weissman, I. L. & Hood, L. (1980) *Nature* (*London*) **283**, 733–739
Dunnick, W., Rabbitts, T. H. & Milstein, C. (1980) *Nature* (*London*) **286**, 669–675
Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. (1980*a*) **19**, 981–998
Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. & Hood, L. (1980*b*) *Cell* **20**, 313–319
Ellison, J. & Hood, L. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1984–1988
Erikson, J., Martinis, J. & Croce, C. M. (1981) Nature (*London*) **294**, 173–175
Flanagan, J. G. & Rabbitts, T. H. (1982*a*) *EMBO J.* **1**, 655–660
Flanagan, J. G. & Rabbitts, T. H. (1982*b*) *Nature* (*London*) **300**, 709–713
Hobart, M. J., Rabbitts, T. H., Goodfellow, P. N., Solomon, E., Chambers, S., Spurr, N. & Povey, S. (1981) *Ann. Hum. Genet.* **45**, 331–335
Honjo, T. & Kataoka, T. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2140–2144
Hozumi, N. & Tonegawa, S. (1976) *Proc. Natl. Sci. U.S.A.* **73**, 3628–3632
Karpas, A., Hayhoe, F. G. J., Greenberger, J. S., Barker, C. R., Cawley, J. C., Lowenthal, R. M. & Moloney, W. C. (1977) *Leukaemia Res.* **1**, 35–49
Kataoka, T., Kawakami, T., Takahashi, N. & Honjo, T. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 919–923
Kirsch, I. R., Morton, C. C., Nakahara, K. & Leder, P. (1982) *Science* **216**, 301–303
Klein, E., Klein, G., Nadkarni, J. S., Nadkarni, J. J., Wigzell, H. & Clifford, P. (1968) *Cancer Res.* **28**, 1300–1310
Klein, G. (1981) *Nature* (*London*) **294**, 313–318
Knapp, W., Bolhuis, R. L. H., Rädl, J. & Hijmars, W. (1973) *J. Immunol.* **111**, 1295–1298
Krawinkel, U. & Rabbitts, T. H. (1982) *EMBO J.* **1**, 403–407
Krumlauf, R., Jeanpierre, M. & Young, B. D. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2971–2975
Kurosawa, Y. & Tonegawa, S. (1982) *J. Exp. Med.* **155**, 201–218
McBride, O. W., Hieter, P. A., Hollis, G. F., Swan, D., Otey, M. C. & Leder, P. (1982) *J. Exp. Med.* **155**, 1480–1490
Malcolm, S., Barton, P., Murphy, C., Ferguson-Smith, M. A., Bentley, D. L. & Rabbitts, T. H. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4957–4961
Marcu, K. B., Lang, R. B., Stanton, L. W. & Harris, L. J. (1982) *Nature* (*London*) **298**, 87–89
Mather, E. & Perry, R. P. (1981) *Nucleic Acids Res.* **9**, 6855–6867
Matthyssens, G. & Rabbitts, T. H. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6561–6565
Max, E. E., Seidman, J. G. & Leder, P. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3450–3454
Milstein, C., Burrone, O. R., Dunnick, W., Milstein, C. P. & Rabbitts, T. H. (1981) in *Mechanisms of Lymphocyte Activation* (Resch & Kirchner, eds.), pp. 63–78, Elsevier, North-Holland, Amsterdam
Minowada, J., Ohnuma, T. & Moore, G. E. (1972) *J. Nat. Cancer Inst.* **49**, 891–895
Pernis, B., Forni, L. & Luzzatti, A. L. (1976) *Cold Spring Harbor Symp. Quant. Biol.* **41**, 175–183
Rabbitts, T. H. (1978) *Nature* (*London*) **275**, 291–296
Rabbitts, T. H. & Forster, A. (1978) *Cell* **13**, 319–327
Rabbitts, T. H., Forster, A., Dunnick, W. & Bentley, D. L. (1980*a*) *Nature* (London) **283**, 351–356
Rabbitts, T. H., Hamlyn, P. H., Matthyssens, G. & Roe, B. A. (1980*b*) *Can. J. Biochem.* **58**, 176–187
Rabbitts, T. H., Forster, A. & Milstein, C. P. (1981) *Nucleic Acids Res.* **9**, 4509–4524
Ravetch, J. V., Siebenlist, U., Korsmeyer, S., Waldmann, T. & Leder, P. (1981) *Cell* **27**, 583–591
Rowley, J. D. (1982) *Science* **216**, 749–751
Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. (1979) *Nature* (*London*) **280**, 288–294
Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980) *Nature* (*London*) **286**, 676–680
Secher, D. S., Cotton, R. G. H. & Milstein, C. (1973) *FEBS Lett.* **37**, 311–316
Siebenlist, U., Ravetch, J. V., Korsmeyer, S., Waldmann, T. & Leder, P. (1981) *Nature* (*London*) **294**, 631–635
Sledge, C., Fair, D. S., Black, B., Krueger, R. G. & Lood, L. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 923–927
Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517
Takahashi, N., Ueda, S., Obata, M., Nikaido, T., Nakai, S. & Honjo, T. (1982) *Cell* **29**, 671–679
Wang, A. C., Wilson, S. K., Hopper, J. E., Fudenberg, H. H. & Nisonoff, A. (1970) *Proc. Natl. Acad. U.S.A.* **66**, 337–343

Attachment E

J. Mol. Biol. (1997) 270, 587–597

JMB

Sequence of the Human Immunoglobulin Diversity (D) Segment Locus: A Systematic Analysis Provides No Evidence for the Use of DIR Segments, Inverted D Segments, "Minor" D Segments or D-D Recombination

Simon J. Corbett[1]*, Ian M. Tomlinson[1], Erik L. L. Sonnhammer[2] David Buck[2] and Greg Winter[1,3]

[1]*MRC Centre For Protein Engineering, Hills Road Cambridge CB2 2QH, UK*

[2]*The Sanger Centre Hinxton Hall, Hinxton, Cambs CB10 1RQ, UK*

[3]*MRC Laboratory Of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*

*Corresponding author

We have determined the complete nucleotide sequence of the human immunoglobulin D segment locus on chromosome 14q32.3 and identified a total of 27 D segments, of which nine are new. Comparison with a database of rearranged heavy chain sequences indicates that the human antibody repertoire is created by VDJ recombination involving 25 of these 27 D segments, extensive processing at the V-D and D-J junctions and use of multiple reading frames. We could find no evidence for the proposed use of DIR segments, inverted D segments, "minor" D segments or D-D recombination. Conventional VDJ recombination, which obeys the 12/23 rule, is therefore sufficient to explain the wealth of lengths and sequences for the third hypervariable loop of human heavy chains.

© 1997 Academic Press Limited

Keywords: diversity; D segment; germline; immunoglobulin; recombination

Introduction

Antigen binding by antibodies is mediated by six polypeptide loops: three from the heavy chain variable domain (H1, H2 and H3) and three from the light chain variable domain (L1, L2 and L3). Although all six loops are variable in sequence and length, the H3 loop, which is located at the centre of the antigen binding site and makes more contacts with antigen than any other loop (Wilson & Stanfield, 1994), is by far the most diverse (Kabat *et al.*, 1991). Diversity in this region is generated by VDJ recombination of three sets of germline gene segments: variable (V_H), diversity (D) and joining (J_H) (Tonegawa, 1983). Joining is directed by specific recombination signal sequences (RSS), which flank the coding sequence of every germline gene segment. Each RSS consists of a conserved palindromic heptamer sequence separated from a conserved nonamer sequence by a 12 or 23 bp spacer. Only segments with differently sized spacers may be joined according to the "12/23 rule". Nucleotides are often removed at the site of recombination and a variable number of non-templated N-nucleotides added by terminal deoxynucleotidyl transferase (TdT; Alt & Baltimore, 1982; Komori *et al.*, 1993). Thus, the sequences of the D segments and the way in which they join to the V_H and J_H segments are major factors in determining antigen specificity.

The human germline V_H and J_H segments have now been completely mapped and sequenced. There are approximately 51 functional V_H segments, depending on the haplotype (Tomlinson *et al.*, 1992; Cook & Tomlinson, 1995), and six functional J_H segments (Ravetch *et al.*, 1981). In contrast, the total number of functional D segments and their sequences are unknown. Indeed, of the estimated 30 segments on chromosome 14 (the "major" D segment locus), only 18 have been sequenced (Ravetch *et al.*, 1981; Siebenlist *et al.*, 1981; Buluwela *et al.*, 1988; Ichihara *et al.*, 1988). In addition, it is not completely understood how these D segments are incorporated into the H3 region. Conventional VDJ recombination involves the joining of the 3′ RSS of the D segment with the 5′ RSS of the J_H segment, followed by the joining of the 5′ RSS of the D segment with the 3′ RSS of the V_H

Abbreviations used: V_H, heavy chain variable gene segment; D, heavy chain diversity gene segment; J_H, heavy chain joining gene segment; RSS, recombination signal sequence; TdT, terminal deoxynucleotidyl transferase.

0022–2836/97/290587–11 $25.00/0/mb971141

© 1997 Academic Press Limited

segment. However, it has also been suggested that D segments may use the 5′ instead of the 3′ RSS to recombine to a J_H segment (in this case, the D segment would be inverted; Gellert, 1992; Tuaillon *et al.*, 1995) and that D segments may recombine with each other (D-D recombination, flouting the 12/23 rule; Sanz, 1991; Yamada *et al.*, 1991; Brezinschek *et al.*, 1995). It has also been suggested that additional elements may be incorporated into the H3 loop. These include the DIR segments, longer sequences interspersed amongst the functional D segments that are flanked by multiple 12 and 23 bp spacer RSS (Ichihara *et al.*, 1988; Sanz *et al.*, 1994) and the "minor" D segments (a cluster of D segments on chromosome 15; Matsuda *et al.*, 1988, 1990; Nagaoka *et al.*, 1994; Tomlinson *et al.*, 1994). Since the evidence for these unconventional mechanisms is based on short sequence homologies with many mismatches and insertions/deletions (Sanz, 1991; Yamada *et al.*, 1991) it is difficult to verify their existence without a complete knowledge of the sequences of all the germline D segments.

To establish the precise mechanism involved in VDJ recombination, we therefore determined the complete nucleotide sequence of the major D segment locus on chromosome 14q32.3. We then used a systematic and quantitative method of scoring sequence homologies with somatically rearranged heavy chains to identify all the germline D segments (and any other sequences) that contribute to H3 diversity.

Results

Complete nucleotide sequence of the D segment locus

A shotgun sequencing strategy (Wilson *et al.*, 1994) was used to sequence three cosmids (COS23, COS21 and COS24) that form a contig covering the entire D segment locus on chromosome 14q32.3 (Buluwela *et al.*, 1988). The nucleotide sequence of 92,588 bp (EMBL data library accession number X97051) was completely determined to high accuracy from a *Hin*dIII restriction site 13 kb upstream of the V_H6 segment to the beginning of the μ switch region 2.7 kb downstream of the J_H segments (Figure 1). The distance from V_H6 to J_H1 is therefore 74 kb.

There are 27 D segments between V_H6 and the J_H segments

In all, 27 conventional D segments were identified that have a short potential coding sequence (11 to 37 bp) flanked by two 12 bp spacer RSS (Figure 2). These can be grouped into seven families based on sequence homology. Six of these families (D_M, D_{LR}, D_{XP}, D_A, D_K and D_N; Ichihara *et al.*, 1988) have at least four members, whilst the seventh is the unique D_{HQ52} segment. Within each family, the RSS are relatively conserved (consistent with recognition by the recombinase), whilst the coding sequences are more divergent (consistent with providing diversity for the H3 loop). In addition, one DIR segment was found immediately

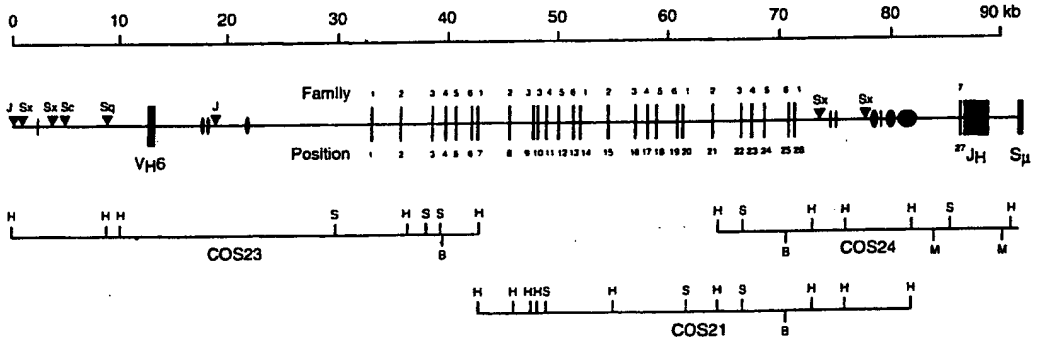


Figure 1. A scale map of the human immunoglobulin D segment locus based on the nucleotide sequences of the three cosmid clones, COS23, COS21 and COS24 (indicated below the map with the following restriction sites; H, *Hin*dIII; S, *Sfi*I; B, *Bss*HII; M, *Mlu*I). COS23 and COS21 have previously been shown to juxtapose (Buluwela & Rabbitts, 1988). D segments are represented by thin vertical lines, *Alu* elements by triangles and L1 elements by ovoids (the width of which indicates the approximate size of the element). The *Alu* sub-families are shown above each element. Where definite sub-family allocations for L1 elements could be made, all were to pre-primate radiation, mammalian-wide families. The V_H6 segment, the J_H segments and the S_μ region are shown as black rectangles. D segments are named according to the convention originally established for the human V_H segments (Shin *et al.*, 1991; Cook *et al.*, 1994). This introduces a two-number system, where the first number indicates the family (shown above each segment in the Figure) and the second number indicates the relative position of the segment in the locus from V_H to J_H (shown below each segment in the Figure). The new family assignments are therefore as follows: D_M becomes D1, D_{LR} becomes D2, D_{XP} becomes D3, D_A becomes D4, D_K becomes D5, D_N becomes D6 and D_{HQ52} becomes D7-27. The new names for all previously sequenced D segments are given in Figure 2.

New Name	Old Name	Nonamer	12 bp Spacer	Heptamer	D Segment Coding Sequence	Heptamer	12 bp Spacer	Nonamer
D1-1		AGATTTTGA	ACAGCCCCGAGT	CACGGTG	GGTACAACTGGAACGAC---	CACCGTG	AGAAAAACTGTG	TCCAAAACT
D1-7	D_{M1}[1]	G.....	.....	...A...	....T.....T..---	...T...	.....G..TC.	.....G
D1-14	D_{M2}[1]	G.....T..	.....	...A.T.	....T...C.....C..---	...T..C	....T.G..AC.	..A.....
D1-20		G.....	.....	...A...	....T.....---	.....	.....	.....
D1-26		G.....	.....	.....	....T.GTG...G.T..TAC	...T...	.....G..A..	.....
D2-2	(D_4[2], D_{4-h}[3]	GGATTTTGT	GGGGGGTCGTGT	CACTGTG	AGGATATTGTAGTAGTACCAGCTGCTATAGCC	CACAGTG	ACACAGCCCCAT	TCCCAAAGC
D2-8	D_{LR1}[1] (D_1[2])	.....	.....	.....	.....C..A.GGTGTA.....	.....	.....	.....
D2-15	D_2[2]	.....	.....	.....	.....G..GGT.....CT..	.....	.....A.....	.....
D2-21	(D_3[2])	.....	.....T.....	.....	..C.....G..G..G---A.....T..	.....	.....A.....	...T.....
D3-3	D_{XP4}[4] (D23/74)	GGTTTGGGG	TGAGGTCTGTGT	CACTGTG	GTATTACGATT---TTTGGAGTGGTTAT---TATACC	CACAGTG	TCACAGAGTCCA	TCAAAAACC
D3-9	D_{XP1}[4], D21/0.5[4]	.....A.AA	.....	.....	.....A---...T..C.....---....A.	.....	.....	.....
D3-10	$D_{XP'1}$[4] (D21/74)	.....	.....	.....	.....T..G---G..C.G.GA.....---....A.	.....	.....	.....
D3-16	(D21/10[4])	.....AA.	.....	.....	.....T....ACG.....G.GA.....CGT.....	.....CA	.....CG.....	...G.....
D3-22	D21/9[4]	.....AA.	.....	.....	.....T..G---A.A.T.....---..CTA.	.....	.....	.....T
D4-4	D_{A4}[1]	GCTTTTTGT	GAAGGGTCCTCC	TACTGTG	TGACTAC---AGTAACTAC	CACAGTG	ATGAACCCAGCA	GCAAAAACT
D4-11	D_{A1}[1]	.....	.....C.....	.G.....	.....---.....	..T....	.....TG	.....
D4-17		.....	.....CC.....	.....	.....---G..G.....	.....	.....A.T....	.....
D4-23		.....	.....C.....	.G.....	.....GGTG.....C.	.....	.....A.....	.....
D5-5	D_{K4}[1]	GGTTATTGT	CAGGGGGTGTCA	GACTGTG	GTGGATACAGCTA---TGGTTAC	CACAGTG	GTGCTGCCCATA	GCAGCAACC
D5-12	D_{K1}[1]	.....	....C.A.....	.....	.....T..TGGCTAC.A....	.....	....C.....	.....
D5-18		.....	.....	.....	.....---.....	.....	.....	.....
D5-24		.....	.....C..	.G.C...	..A..G.TG....---CAA....	.....	....C.....	.....
D6-6	(D_{N4}[1])	AGTTTCTGA	AGGTGTCTGTGT	CACAGTG	GAGTATAGCA---GCTCGTCC	CACAGTG	ACACTGGCCAGG	CCAGAAACC
D6-13	D_{N1}[1]	G.....	.....	.....	.G.....GCA...G..A.	.....	.....A..CA.	.....
D6-19		G.....	..C.....A.	.....	.G.....GTG...G..A.	.....	.....	.....
D6-25		G.....	..C.....	C	.G.....---..GGC.A.	....A..	.....G.G....	A.....
D7-27	(D_{HQ52}[5])	GGTTTTGGC	TGAGCTGAGAAC	CACTGTG	CTAACTGGGGA	CACAGTG	ATTGGCAGCTCT	ACAAAAACC

Figure 2. Nucleotide sequences of the human D segments on chromosome 14q32.3. Sequences are aligned in family groups with dots to indicate nucleotide homology and dashes to indicate pad characters (inserted to maximise identity). For each D segment the RSS heptamers and nonamers are boxed. Nucleotides that would be expected to prevent or severely restrict recombination (Gellert, 1992; Akamatsu *et al.*, 1994) are circled. Names of previously published sequences are given next to their corresponding D segments. References are [1]Ichihara *et al.* (1988); [2]Siebenlist *et al.* (1981); [3]Zong *et al.* (1988); [4]Buluwela *et al.* (1988); [5]Ravetch *et al.* (1981). Note that D_2, D_3 and D_4 were later renamed D_{LR2}, D_{LR3} and D_{LR4} (Ichihara *et al.*, 1988). Sequence names in parentheses have a number of nucleotide substitutions (underlined positions) compared to our sequences. These are: D_4 (allele of D2-2) with A → G at the second underlined position; D_{4-h} (allele of D2-2) with A → T and A → G at the first and second underlined positions, respectively; D_1 (allele of D2-8) with AA → GG, D_3 (allele of D2-21) with T → C and C → T at the first and second underlined positions, respectively; D_{N4} (allele of D6-6) with A → C; and D_{HQ52} (allele of D7-27) with G → T. The previously published sequences of D23/7 (from cosmid COS23), D21/7 (from cosmid COS21), D21/10 (from cosmid COS21) (Buluwela *et al.*, 1988) are incorrect and are therefore superseded by the sequences of D3-3, D3-10 and D3-16 shown here. Note that the coding sequence of D4-4 is the same as D4-11 and that of D5-5 is the same as D5-18.

upstream of each of the five D_M segments. As all the D segments in the major locus on chromosome 14 have now been identified, we propose a new naming convention (similar to the one used for the human V_H segments) that for each D segment indicates its family and its relative position in the locus (see the legend to Figure 1 and Figure 2). (DIR segments were omitted from this scheme as there is no evidence that they contribute to the functional repertoire, see below.)

Comparison of our and previously published D segment sequences indicates limited allelic polymorphism (see Figure 2). The sequences of all human germline D segments can be obtained from the V BASE directory of human immunoglobulin genes (available on the World Wide Web: http://www.mrc-cpe.cam.ac.uk/imt-doc/vbase-homepage.html or on disk from I.M.T.).

Organisation and evolution of the D segment locus

The complete nucleotide sequence of the D segment locus on chromosome 14 confirms that the major D segments are arranged in four 9.5 kb tandem repeat units, with the same order of D segment families in each cluster (Figure 3a; Siebenlist *et al.*, 1981; Buluwela *et al.*, 1988; Ichihara *et al.*, 1988; Zong *et al.*, 1988). The high degree of sequence homology is illustrated by a two-dimensional dot matrix plot of this region (Figure 3b). In addition to the four 9.5 kb repeats there is a 2.8 kb

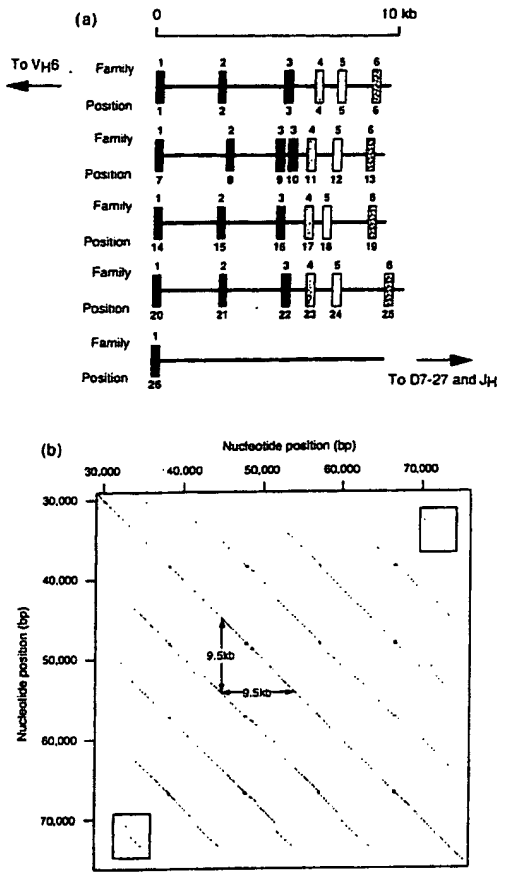


Figure 3. Repeat structure of the central D segment cluster. a, A representation of the central D segment cluster containing 26 segments (drawn to scale). The six families are represented by different shading. The family is shown above and the position is shown below each segment. b, Two-dimensional dot plot created with the program Dotter (Sonnhammer & Durbin, 1995) from nucleotides 29,000 to 76,000 (as shown in Figure 1) which encompasses the central D segment cluster. The shortest distance between repeats (indicated by arrows) is, on average, 9.5 kb. In addition to four 9.5 kb repeats there is a region of homology of 2.8 kb shown boxed at the bottom left and top right of the diagram. Breaks in the continuity of the diagonals indicate small-scale insertions/deletions from one repeat unit to the next.

region homologous to one end of a full repeat. This contains an extra D1 segment (boxed in Figure 3b).

Although the locus is G + C-rich (54.9%) and *Alu* repeat elements tend to concentrate in G + C-rich regions of the genome (Korenberg & Rykowski, 1988), the density of *Alu* repeats (0.05/kb) is lower than the lowest previously reported for a large human sequence (0.1/kb; Legouis *et al.*, 1991; Whitfield *et al.*, 1995). Indeed, the central D segment cluster (D1-1 to D1-26) lacks *Alu* and L1 elements altogether (Figure 1). This suggests that the primordial locus lacked *Alu*/L1 elements and that the flanking *Alu*/L1 elements were spread apart by repeated duplication of the D segments. If true, this indicates that the expansion of the central D segment locus occurred within the last 33 million years (when the youngest *Alu* elements, *Alu*-S, were dispersed throughout the genome; Zietkiewicz *et al.*, 1994). This is consistent with the different organisation and family classification of the central D segment cluster in mice (Ichihara *et al.*, 1989; Feeney & Riblet, 1993). The only D segment in humans (D7-27/D_{HQ52}, Figure 1) related to a mouse D segment (D_{Q52}, Ravetch *et al.*, 1981) is located in a similar position immediately upstream of the J_H segments, indicating that this segment was present prior to the divergence of the human and mouse lineages.

Alignment of rearranged heavy chain H3 sequences

To assign somatically rearranged sequences to their germline V, D and J counterparts, we used the BLAST algorithm (Altschul *et al.*, 1990). BLAST does not insert gaps to maximise homology between query and target sequences and produces a numerical score for the best alignment. The BLAST algorithm was incorporated into a UNIX-based sequence alignment package called Germline Query (GQ), which takes the sequence of a rearranged antibody gene and systematically searches a directory of germline gene segments (V BASE) for the best matches to V and J segments (Figure 4). With heavy chain genes, GQ masks the region of the query sequence that aligns to V_H and J_H and then searches for D segments in the intervening H3 sequence.

In all, 893 different heavy chain sequences with identifiable V_H and J_H sequences were extracted from a large database of rearranged human heavy chain sequences (Tomlinson *et al.*, 1996) and used as the source of query sequences for GQ. We also compiled a database of target D sequences, corresponding to the nucleotide sequences (both strands) of the 27 germline D segments and their alleles (see Figure 2), the five D segment sequences from chromosome 15 (Matsuda *et al.*, 1988, 1990; Nagaoka *et al.*, 1994; Tomlinson *et al.*, 1994) and the five DIR segments (from this study).

In previous studies, the criteria for assigning H3 sequences to their germline D segments appear to have been far too lenient. For example, others (Mortari *et al.*, 1993; Brezinschek *et al.*, 1995) have taken five nucleotides of identity or six nucleotides with one mismatch as their minimum cut-off for identifying a germline sequence. In the complete D segment database of 3978 bp (see above) even a random sequence for H3 is likely find a D segment match according to these criteria (there are only 1024 possible sequence permutations of five

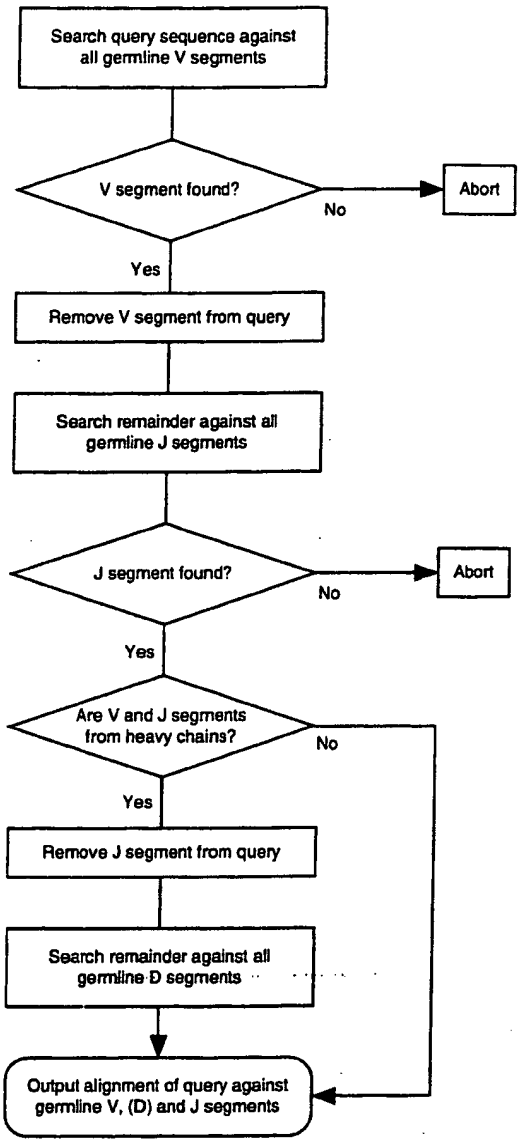


Figure 4. Flow-chart for the GQ alignment software.

nucleotides). Thus, many of these previous alignments are probably incorrect.

As a control for D segment assignment, we therefore generated 28,000 "mock" heavy chains with random H3 sequences of varying lengths sandwiched between genuine V_H and J_H sequences (H3 length ranged from five to 70 nucleotides in intervals of five nucleotides, with 2000 sequences of each length). The scores produced by aligning these sequences to the germline D segment database provided a "threshold of confidence" for alignment of real heavy chain H3 sequences. Our criteria were stringent: if the score for an H3 alignment rose above the 99th percentile score produced by mock sequences of the same length the alignment was considered plausible. Even with this high threshold of confidence we would expect 1 in 100 alignments to be due to chance and therefore be incorrect. Using this approach, we found that at least ten consecutive nucleotides of identity are generally required to confidently assign a D segment, although the precise threshold depends on the H3 length. To validate our criteria we compared the 893 somatically rearranged heavy chain sequences with the germline D segment database. Reassuringly, the scores for the majority of H3 sequences aligned to D segments from the seven families rise above the 99% threshold (Figure 5a).

No evidence for the use of DIR segments, inverted D segments, "minor" D segments or D-D recombination

The contribution of DIR segments to the heavy chain repertoire was determined by comparing the scores of somatic H3 sequences assigned to DIR segments with our threshold of confidence (Figure 5b). Only six H3 alignments to DIR segments rise above the 99% confidence level, providing little evidence for the use of DIR segments. The DIR segments were therefore removed from the germline D segment database for subsequent alignments.

We then checked for evidence of D segments that rearrange by inversion. Scores produced by alignment to the positive (defined as the one that runs 5' to 3' from V_H to J_H) and negative DNA strands (defined as the one that runs 5' to 3' from J_H to V_H) are shown in Figure 5c and d, respectively. Only seven alignments to the negative strand rise above the 99% confidence threshold, in marked contrast to positive strand alignments, providing little evidence for inversion of D segments during recombination. Subsequent H3 alignments were therefore performed using only the positive strand.

Next, we looked at the contribution of the minor D segments on chromosome 15. The scores produced by alignment of H3 sequences to D segments on chromosomes 14 and 15 are shown in Figure 5e and f, respectively. Only six alignments for the chromosome 15 segments rise above the 99% confidence level. Given that the V_H sequences associated with these H3 alignments are highly mutated (>7.9% nucleotide differences), these probably represent mutated H3 sequences derived from major D segments. Hence, there is little evidence for use of minor D segments. The minor D segments were therefore removed from the germline D segment database for subsequent alignments.

We then searched for evidence of multiple D segment alignments. Any H3 sequence that remained after confident assignment of a first D segment (Figure 5g) was again assigned to the

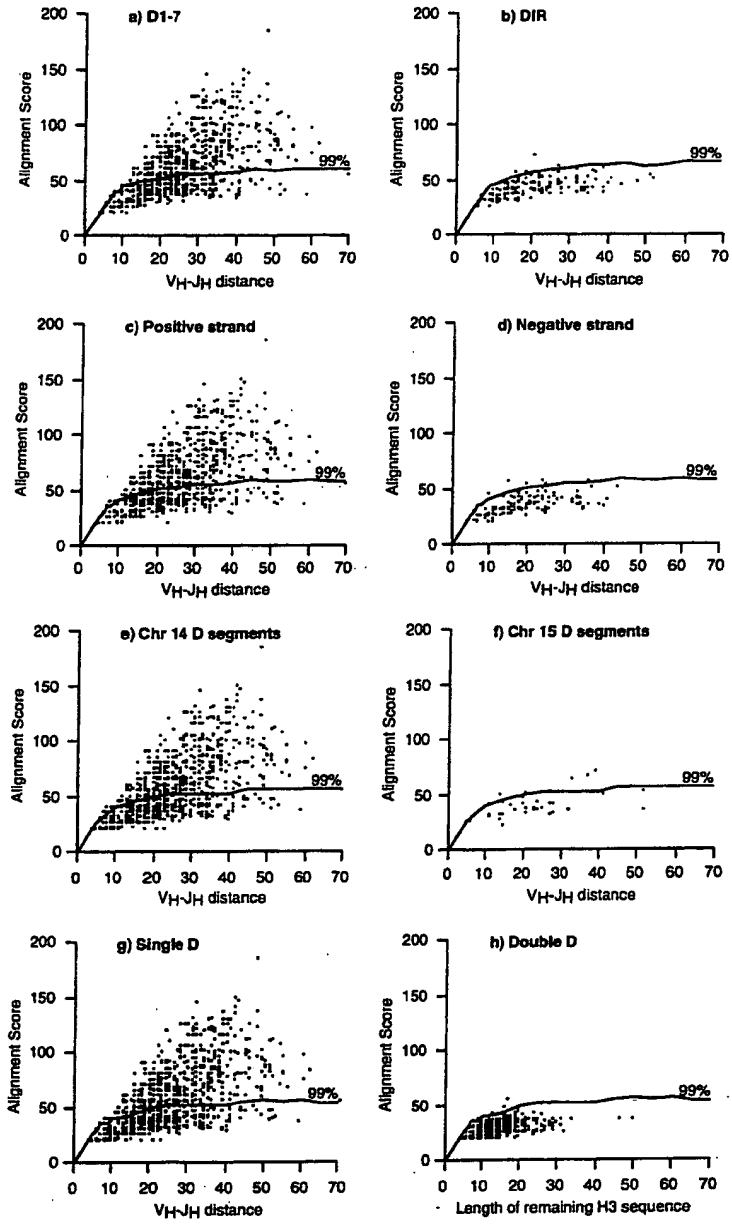


Figure 5. Systematic assignment of somatically rearranged H3 sequences to germline D segments. The scores produced by the best alignment found by GQ when 893 H3 sequences were searched against the germline D segment database are plotted as dots according to the length of the nucleotide sequence used for D segment alignment (the region between the end of V_H homology and the beginning of J_H homology: the V_H-J_H distance). Black lines show the 99th percentile score produced by random sequences aligned to the same catagory of D segments (see the text). For example, the black line in b corresponds to the 99th percentile of random H3 sequences assigned to DIR segments. Note that where two or more alignments have the same score and V_H-J_H distance, a single dot is shown. a, Alignments to D1-7 family members when H3 sequences were screened against a germline D segment database containing both strands of the 27 D segments on chromosome 14 (including their alleles), the five D segment sequences from chromosome 15 and the five DIR segments (from this study). Out of 739 alignments, 399 are above the 99% confidence level (the rest of the 893 rearranged H3 sequences align to DIR segments, see b, below). b, Alignments to DIR segments when H3 sequences were screened against the same germline D segment database used in a. Only six out of 154 alignments are above the 99% confidence level (the rest of the 893 rearranged H3 sequences align to D1-7 segments, see a, above). c, Alignments to the positive DNA strand (defined as the one that runs 5′ to 3′ from V_H to J_H) when H3 sequences were screened against a germline D segment database containing both strands of the 27 D segments on chromosome 14 (including their alleles) and the five D segment sequences from chromosome 15. Out of 766 alignments, 421 are above the 99% confidence level (the rest of the 893 rearranged H3 sequences align to the negative DNA strand, see d, below). d, Alignments to the negative DNA strand (defined as the one that runs 5′ to 3′ from J_H to V_H) when H3 sequences were screened against the same germline D segment database used in c. Only seven out of 127 alignments are above the 99% confidence level (the rest of the 893 rearranged H3 sequences align to the positive DNA strand, see c, above). e, Alignments to major D segments from chromosome 14 when H3 sequences were screened against a D segment database containing the 5′ to 3′ (V_H to J_H) strand of the 27 D segments on chromosome 14 (including their alleles) and the five D segments sequences from chromosome 15. Out of 860 alignments, 454 are above the 99% confidence level (the rest of the 893 rearranged H3 sequences align to minor D segments, see f, below). f, Alignments to minor D segments from chromosome 15 when H3 sequences were screened against the same germline D segment database used in e. Six out of 33 alignments are above the 99% confidence level (the rest of the 893 rearranged H3 sequences align to major D segments, see e, above). g, Alignments to a first D segment when H3 sequences were screened against a D segment database containing the 5′ to 3′ (V_H to J_H) strand of the 27 D segments on chromosome 14 (including their alleles). Out of 893 alignments, 451 are above the 99% confidence level. h, Alignments to a second D segment when the remaining H3 regions 5′ and 3′ of the first D segment alignment (see above) were screened against the same germline D segment database used in g. Only four out of 821 alignments are above the 99% confidence level.

Table 1. Nucleotide composition of H3 sequences

	%G	%C	%A	%T	Average nucleotide length
Total N-nucleotide addition	32.6	28.8	19.1	19.5	13.6
V_H end N-nucleotide addition	32.0	28.9	18.6	20.5	7.3
J_H end N-nucleotide addition	33.3	28.7	19.5	18.4	6.3
No D segment in H3	32.4	23.9	20.6	23.1	22.3
D segments	27.5	16.7	25.9	29.9	

germline D segment database. Only four alignments are above the 99% confidence threshold, providing little evidence for multiple D segments (Figure 5h).

Finally, to check for other germline elements that contribute to H3, sequences flanked by unmutated V_H and J_H segments that could not be confidently assigned to one of the major D segments were realigned to the complete nucleotide sequence of the D locus. From the 31 sequences examined, we were unable to identify any germline alignments above the 99% confidence threshold.

We conclude that the 27 major D segments on chromosome 14 (and their alleles) are the only germline elements that contribute to H3 diversity. In addition, there is little evidence for D segment inversion or D-D recombination.

N-nucleotide addition creates significant H3 diversity

Any "extra" H3 sequence remaining after the confident assignment of V_H, D and J_H segments must be due to N or P-nucleotide addition (Alt & Baltimore, 1982; Lafaille *et al.*, 1989). Since P-nucleotide addition is not thought to occur at nucleolytically processed segment ends (Lafaille *et al.*, 1989) and more than 96% of our 893 somatically rearranged sequences do appear to be processed (data not shown), it seems likely that P-nucleotide

Table 2. D segment and reading frame use in 893 heavy chain sequences

	% of 893 H3	Stop	%	Hydrophilic	%	Hydrophobic	%
D1-1	0.56	VQLER	0	YNWND	60.0	GTTGT	40.0
D1-7	0.45	V*LEL	0	YNWNY	75.0	GITGT	25.0
D1-14	0	V*PEP	0	YNRNH	0	GITGT	0
D1-20	0.67	V*LER	0	YNWND	66.7	GITGT	33.3
D1-26	1.68	V*WELL	0	YSGSYY	46.7	GIVGAT	53.3
D2-2	3.70	RIL**YQLLY	6.0	GYCSSTSCYT	66.7	DIVVVPAAI	27.3
D2-8	0.78	RILY*WCMLY	0	GYCTNGVCYT	71.4	DIVLMVYAI	28.6
D2-15	2.35	RIL*WW*LLL	4.7	GYCSGGSCYS	66.7	DIVVVVAAT	28.6
D2-21	1.23	SILWW*LLF	0	AYCGGDCYS	54.6	HIVVVTAI	45.4
D3-3	4.82	VLRFLEWLLY	18.6	YYDFWSGYYT	58.1	ITIFGVVII	23.3
D3-9	2.13	VLRYFDWLL*	36.8	YYDILTGYYN	63.2	ITIF*LVII	0
D3-10	8.06	VLLWFGELL*	11.1	YYYGSGSYYN	52.8	ITMVRGVII	36.1
D3-16	1.01	VL*LRLGELSLY	0	YYDYVWGSYRYT	66.7	IMITFGGVIVI	33.3
D3-22	3.81	VLL***WLLL	2.9	YYYDSSGYYY	88.3	ITMIVVVIT	8.8
D4-4	0.34	*LQ*L	0	DYSNY	83.3	TTVT	16.7
D4-11	0.33	*LQ*L	0	DYSNY	83.3	TTVT	16.7
D4-17	2.35	*LR*L	0	DYGDY	61.9	TTVT	38.1
D4-23	1.12	*LRW*L	0	DYGGNS	50.0	TTVVT	50.0
D5-5	1.18	WIQLWL	4.8	GYSYGY	61.9	VDTAMV	33.3
D5-12	1.57	WI*WLRL	0	GYSGYDY	71.4	VDIVATI	28.6
D5-18	1.17	WIQLWL	4.8	GYSYGY	61.9	VDTAMV	33.3
D5-24	0.78	*RWLQL	0	RDGYNY	57.1	VEMATI	42.9
D6-6	1.34	V*QLV	8.3	EYSSSS	33.4	SIAAR	58.3
D6-13	3.47	V*QQLV	8.8	GYSSSWY	61.3	GIAAAG	29.0
D6-19	4.70	V*QWLV	11.9	GYSSGWY	42.9	GIAVAG	45.2
D6-25	0	V*QRL	0	GYSSGY	0	GIAAA	0
D7-27	0.90	*LG	0	NWG	25.0	LTG	75.0
No D	49.5						

The first column shows the percentage of the 893 somatically rearranged H3 sequences that can be confidently aligned to each D segment (alignments that exceed the 99% confidence threshold). For the pairs of identical D segments, D4-4/D4-11 and D5-5/D5-18, the percentages of H3 sequences that align to them are divided equally. The rest of the Table shows the amino acid translation of each reading frame of all 27 D segments in single-letter code. The percentage of H3 sequences using each reading frame of any given D segment with respect to J_H, is shown alongside the amino acid translation. Stop codons are shown as asterisks (*). Amino acid sequences in bold type were identified amongst the 893 H3 sequences. Reading frames are defined according to the presence of stop codons, hydrophilic or hydrophobic amino acid residues (see the text). Note that for 49.5% of the 893 rearranged H3 sequences we were unable to confidently assign a D segment (No D).

addition plays a minor role in H3 diversification. In addition, the nucleotide composition of the extra H3 sequence (Table 1) is similar to that ascribed to N-nucleotide addition by terminal deoxynucleotidyl transferase (TdT; Basu *et al.*, 1983; Gauss & Lieber, 1996). In contrast to a recent report on N-nucleotide addition in T-cell receptor β chains (Rowen *et al.*, 1996), there is no difference between the nucleotide composition of N-nucleotides 5′ and 3′ of the D segment (Table 1), suggesting that in B-cells, TdT acts at all segment termini rather than predominantly from D segments as seems to be the case in T-cells.

Although some of the 442 H3 sequences for which a germline D segment could not be confidently assigned may be highly mutated versions of known D segments, most are G + C-rich, suggesting that they may be mainly generated by the activity of TdT (for comparison, Table 1 shows the nucleotide composition of the 27 germline D segments from chromosome 14). In these cases, the germline D segment may be acting purely as an anchor for considerable N-nucleotide diversification.

Use of different D segments and D segment reading frames

In total, 451 rearranged H3 sequences in our database can be confidently assigned to one of the 27 D segments on chromosome 14 (Figure 5g). The use of each D segment is summarised in Table 2. As noted in previous studies, the D3, D6 and D2 families predominate (Sanz, 1991; Yamada *et al.*, 1991; Huang *et al.*, 1992; Brezinschek *et al.*, 1995), with nine D segments (D2-2, D2-15, D3-3, D3-9, D3-10, D3-22, D4-17, D6-13 and D6-19) being used more often than would be expected at random. We note that two of the 27 D segments are not seen at all (D1-14 and D6-25): these segments have mutations within their heptamer sequences that would be expected to prevent or severely restrict recombination (Figure 2; Gellert, 1992; Akamatsu *et al.*, 1994). A similar mutation in the heptamer of D4-4 may also prevent its rearrangement (Figure 2), but since its coding sequence is the same as D4-11 it is impossible to determine which of the two segments is used.

For each D segment family there is one reading frame that tends to encode a stop codon, followed by a reading frame that tends to encode glycine residues in conjunction with polar/hydrophilic residues (especially tyrosine and serine), followed by a third reading frame that is normally hydrophobic in character (with many alanine, valine and isoleucine residues). Instead of defining each reading frame relative to the RSS (Ichihara *et al.*, 1988) we therefore define it according to its character (see Table 2). The frequency of use of the three reading frames for each D segment is given in Table 2, together with an amino acid translation of each frame. The presence of a stop codon does not prevent use of this reading frame, as nucleotide loss during VDJ recombination may remove it. The hydrophilic reading frame predominates in those human H3 sequences for which a confident D segment assignment can be made (60% of sequences, Table 2). In contrast to most of the other segments that normally use the hydrophilic reading frame, the D segments D6-6 and D6-19 are most frequently seen in the hydrophobic reading frame (Table 2).

Discussion

By systematic comparison of the complete sequence of the human immunoglobulin D segment locus with a database of rearranged sequences, we found that DIR segments, inverted D segments, minor D segments or D-D joins produce alignments no better than those produced by randomly generated sequences. In marked contrast, we found overwhelming evidence for conventional VDJ recombination involving 25 of the 27 D segments on chromosome 14, as shown by the majority of alignments scoring higher than those produced by randomly generated sequences. A recent PCR-based study also failed to detect any DIR segment rearrangement in developing B-cells (Moore & Meek, 1995). In the absence of any real evidence, we therefore believe that it is unreasonable to propose a role for the use of DIR segments, inverted D segments, "minor" D segments or D-D joins in the human antibody repertoire.

We have also shown that the use of different D segments and D segment reading frames is highly biased (Table 2). It is an inevitable consequence of scoring sequence alignments according to length that shorter D segments will have fewer rearranged counterparts. Nevertheless, other factors clearly influence the pattern of D segment use, as there are numerous examples of short D segments (most notably D6-19 and D4-17) being seen more often than long D segments (see Table 2). In addition, the fact that the use of D3-10 is twice that of D3-3 (which have identical RSS and are the same length) and that of D2-2 is five times that of D2-8 (which also have identical RSS and length) indicates that structural or selective pressures acting at the amino acid level must affect D segment use. This is also probably the major factor that affects reading frame use (which is not subject to any length biases). The most popular reading frame of almost all human D segments (Table 2) frequently encodes serine, tyrosine and glycine residues. This amino acid composition is similar to that seen in the five other antigen binding loops (see V BASE), suggesting that these residues are particularly good at binding antigen and/or for recruiting somatic hypermutation (for a more in depth discussion of D segment use, see Corbett, 1996).

For many of the rearranged sequences it was impossible to identify a source for the nucleotides located between the end of the V_H segment and the

beginning of the J_H segment. On average, these H3 sequences are shorter (22.3 nucleotides, Table 1) than those for which D segments can be confidently assigned (29.6 nucleotides) and are probably the result of either extensive nucleotide loss from D segments together with N-nucleotide addition and/or high levels of somatic mutation (both of which are accepted mechanisms of diversification). Indeed, on the basis of their nucleotide composition "fingerprint" (Table 1) and the average length of N-nucleotide addition in H3 sequences where D segments have been assigned (13.6 nucleotides, Table 1), it appears that these sequences are mainly the result of TdT activity.

Despite the mechanistic constraints and the various biases described above, there is still more sequence and length diversity in the human H3 region than in any other species so far examined (Wu *et al.*, 1993). The mouse (Ichihara *et al.*, 1989; Feeney & Riblet, 1993), chicken (Reynaud *et al.*, 1991) and rabbit (Knight & Crane, 1994) all have fewer germline D segments and fewer families. Furthermore, their D segments are usually (>80%) expressed in the hydrophilic reading frame (Kaartinen & Makela, 1985; Gu *et al.*, 1991; Knight & Crane, 1994; Reynaud *et al.*, 1994), whereas in humans only 60% of antibodies use the hydrophilic reading frame (Table 2). In this way, the use of many different D segments and reading frames in humans by conventional VDJ recombination together with extensive N-nucleotide addition creates a wide range of H3 structures and hence a vast repertoire of antigen binding sites.

Materials and Methods

Shotgun sequencing of cosmids COS21 COS23 and COS24

The shotgun sequencing strategy was as described (Wilson *et al.*, 1994; Baxendale *et al.*, 1995). Briefly, cosmids were fragmented by sonication and randomly subcloned into M13mp18. Greater than sevenfold coverage was achieved by sequencing at least 600 single-stranded M13 DNA templates per cosmid. Single-stranded DNA was prepared from subclones using anti-M13 magnetic beads. Templates were sequenced using the dideoxy chain termination cycle sequencing technique with dye-labelled M13-specific forward primers. Products were then run on an Applied Biosystems 373A stretch-liner sequencer. The sequences of COS21 and COS24 were compiled and assembled with the Staden package (Staden, 1994), whilst COS23 was initially assembled using the phrap algorithm (P. G. Green, unpublished) and was then incorporated into Staden package format by phrap2gap (S. Dear, R. Durbin, R. Mott, G. Marth & D. Hodgson, unpublished) for further manipulation. Sequence problems were resolved using a combination of reverse reads, long reads, oligo walks and Taq terminators. The accuracy of the entire 92,588 bp sequenced is estimated to be 99.98% with the methodology used (Wilson *et al.*, 1994; Whitfield *et al.*, 1995). The EMBL data library accession number for the complete sequence is X97051.

Computer-aided sequence analysis

This was performed as described (Whitfield *et al.*, 1995). Briefly, completed sequence was analysed on a cosmid-by-cosmid basis. Each cosmid was screened for repetitive elements using: the human repeat sequence database assembled by J. Jurka available from the PYTHIA server (email; pythia@anl.gov); a hidden Markov model to identify *Alu* elements; BLASTN to identify L1 and MER/SINE sequences; the programs QUICK-TANDEM and TANDEM to identify tandem repeats. The subfamily allocation of *Alu* and L1 elements was performed using the programs ALUS and CENSOR from the PYTHIA server, respectively. Repetitive sequence was then masked with the character N before BLASTN (B = 1,000,000) and BLASTX (B = 1,000,000, S = 50, M = BLOSUM62-12) were used to screen the remaining non-repetitive sequence against the following databases to identify coding regions and other features of interest: EMBL release 39; SWIR version 8 (which is a Sanger Centre non-redundant compilation of SWISS-PROT31, PIR43 and WORMPEP8); dbEST version 3.5. No other significant potential coding region could be identified with GRAIL II or by searching for CpG islands with the program CPG.

Germline Query (GQ)

GQ is an AWK script that takes an antibody variable region sequence and uses the BLAST algorithm to determine the best germline V, D and J segment alignments in the V BASE directory of human immunoglobulin genes (available on the World Wide Web: http://www.mrc-cpe.cam.ac.uk/imt-doc/vbase-home-page.html or on disk from I.M.T.). The GQ alignment process is summarised in Figure 4. V_H and J_H alignments were accepted if they exceeded 50% and 85% homology, respectively. In addition, V_H alignments were rejected if they ended more than ten nucleotides before the 3' end of the germline segment sequence, and J_H segment alignments had to be a minimum length of 18 nucleotides. The boundaries of segment alignments were defined by a series of contiguous mismatches with boundary conditions set as seven, two and three nucleotides for the V_H, D and J_H regions, respectively. Alignment scores were calculated using a system of +5 for a match and −4 for a mismatch.

Acknowledgements

The authors thank Dr Terry Rabbitts for donating cosmids, Dr Matt Jones for making cosmid libraries, all the members of human sequencing team 31 at the Sanger Centre for their help and advice during the sequencing stage of the project, Dr Gos Micklem for help with sequence analysis and Dr Paul Dear for help with generating random sequences. S.J.C. is supported by an MRC Studentship and the University of Cambridge MB/PhD programme. The Wellcome Trust funds human sequencing and analysis at the Sanger Centre.

References

Akamatsu, Y., Tsurushita, N., Nagawa, F., Matsuoka, M., Okazaki, K., Imai, M. & Sakano, H. (1994).

Essential residues in V(D)J recombination signals. *J. Immunol.*, **153**, 4520–4529.

Alt, F. W. & Baltimore, D. (1982). Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-J_H fusions. *Proc. Natl Acad. Sci. USA*, **79**, 4118–4122.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Basu, M., Hegde, M. V. & Modak, M. J. (1983). Synthesis of compositionally unique DNA by terminal deoxynucleotidyl transferase. *Biochem. Biophys. Res. Commun.* **111**, 1105–1112.

Baxendale, S., Abdulla, S., Elgar, G., Buck, D., Berks, M., Micklem, G., Durbin, R., Bates, G., Brenner, S., Beck, S. & Lehrach, H. (1995). Comparative sequence analysis of the human and pufferfish Huntingdon's disease genes. *Nature Genet.* **10**, 67–76.

Brezinschek, H. P., Brezinschek, R. I. & Lipsky, P. E. (1995). Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. *J. Immunol.* **155**, 190–202.

Buluwela, L. & Rabbitts, T. H. (1988). A V_H gene is located within 95 kb of the human immunoglobulin heavy chain constant region genes. *Eur. J. Immunol.* **18**, 1843–1845.

Buluwela, L., Albertson, D. G., Sherrington, P., Rabbitts, P. H., Spurr, N. & Rabbitts, T. H. (1988). The use of chromosomal translocations to study human immunoglobulin gene organization: mapping D_H segments within 35 kb of the Cμ gene and identification of a new D_H locus. *EMBO J.* **7**, 2003–2010.

Cook, G. P. & Tomlinson, I. M. (1995). The human immunoglobulin V_H repertoire. *Immunol. Today*, **16**, 237–242.

Cook, G. P., Tomlinson, I. M., Walter, G., Riethman, H., Carter, N. P., Buluwela, L., Winter, G. & Rabbitts, T. H. (1994). A map of the human immunoglobulin V_H locus completed by analysis of the telomeric region of chromosome 14q. *Nature Genet.* **7**, 162–168.

Corbett, S. J. (1996). The human immunoglobulin diversity segment repertoire. PhD thesis, University of Cambridge, UK.

Feeney, A. J. & Riblet, R. (1993). D_{ST4}: a new and probably the last, functional D_H gene in the BALB/c mouse. *Immunogenetics*, **37**, 217–221.

Gauss, G. H. & Lieber, M. R. (1996). Mechanistic constraints on diversity in human V(D)J recombination. *Mol. Cell. Biol.* **16**, 258–269.

Gellert, M. (1992). Molecular analysis of V(D)J recombination. *Annu. Rev. Genet.* **22**, 425–446.

Gu, H., Kitamura, D. & Rajewsky, K. (1991). B Cell development regulated by gene rearrangement: arrest of maturation by membrane-bound Dμ protein and selection of D_H element reading frames. *Cell*, **65**, 47–54.

Huang, C., Stewart, A. K., Schwartz, R. S. & Stollar, B. D. (1992). Immunoglobulin heavy chain gene expression in peripheral blood B lymphocytes. *J. Clin. Invest.* **89**, 1331–1343.

Ichihara, Y., Matsuoka, H. & Kurosawa, Y. (1988). Organization of human immunoglobulin heavy chain diversity gene loci. *EMBO J.* **7**, 4141–4150.

Ichihara, Y., Hayashida, H., Miyazawa, S. & Kurosawa, Y. (1989). Only D_{FL16}, D_{SP2}, and D_{Q52} gene families exist in mouse immunoglobulin heavy chain diversity loci, of which D_{FL16} and D_{SP2} originate from the same primordial D_H gene. *Eur. J. Immunol.* **19**, 1849–1854.

Kaartinen, M. & Makela, O. (1985). Reading of D genes in variable frames as a source of antibody diversity. *Immunol. Today*, **6**, 324–327.

Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest*, vols 1, 2 and 3, U.S. Department of Health and Human Services.

Knight, K. L. & Crane, M. A. (1994). Generating the antibody repertoire in rabbit. *Advan. Immunol.* **56**, 179–218.

Komori, T., Okada, A., Stewart, V. & Alt, F. W. (1993). Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science*, **261**, 1171–1175.

Korenberg, J. R. & Rykowski, M. C. (1988). Human genome organisation: Alu, Lines, and the molecular structure of metaphase chromosome bands. *Cell*, **53**, 391–400.

Lafaille, J. J., DeCloux, A., Bonneville, M., Takagaki, Y. & Tonegawa, S. (1989). Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V-(D)-J joining. *Cell*, **59**, 859–870.

Legouis, R., Hardelin, J.-P., Levilliers, J., Claverie, J.-M., Compain, S., Wunderle, V., Millasseau, P., Le Paslier, D., Cohen, D., Caterina, D., Bougueleret, L., Delemarre-Van de Waal, H., Lutfalla, G., Weissenbach, J. & Petit, C. (1991). The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. *Cell*, **67**, 423–435.

Matsuda, F., Lee, K. H., Nakai, S., Sato, T., Kodaira, M., Zong, S. Q., Ohno, H., Fukuhara, S. & Honjo, T. (1988). Dispersed localization of D segments in the human immunoglobulin heavy-chain locus. *EMBO J.* **7**, 1047–1051.

Matsuda, F., Shin, E. K., Hirabayashi, Y., Nagaoka, H., Yoshida, M. C., Zong, S. Q. & Honjo, T. (1990). Organization of variable region segments of the human immunoglobulin heavy chain: duplication of the D5 cluster within the locus and interchromosomal translocation of variable region segments. *EMBO J.* **9**, 2501–2506.

Moore, B. B. & Meek, K. (1995). Recombination potential of the human DIR elements. *J. Immunol.* **154**, 2175–2187.

Mortari, F., Wang, J.-Y. & Schroeder, H. W., Jr (1993). Human cord blood antibody repertoire. *J. Immunol.* **150**, 1348–1357.

Nagaoka, H., Ozawa, K., Matsuda, F., Hayashida, H., Matsumura, R., Haino, M., Shin, E. K., Fukita, Y., Imai, T., Anand, R., Yokoyama, K., Eki, T., Soeda, E. & Honjo, T. (1994). Recent translocation of variable and diversity segments of the human immunoglobulin heavy chain from chromosome 14 to chromosomes 15 and 16. *Genomics*, **22**, 189–197.

Ravetch, J. V., Siebenlist, U., Korsmeyer, S., Waldmann, T. & Leder, P. (1981). Structure of the human immunoglobulin μ locus: characterization of embryonic and rearranged J and D genes. *Cell*, **27**, 583–591.

Reynaud, C.-A., Anquez, V. & Weill, J.-C. (1991). The chicken D locus and its contribution to the immunoglobulin heavy chain repertoire. *Eur. J. Immunol.* **21**, 2661–2670.

Reynaud, C.-A., Bertocci, B., Dahan, A. & Weill, J.-C. (1994). Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. *Advan. Immunol.* **57**, 353–378.

Rowen, L., Koop, B. F. & Hood, L. (1996). The complete 685-kilobase DNA sequence of the human β T cell receptor locus. *Science*, **272**, 1755–1762.

Sanz, I. (1991). Multiple mechanisms participate in the generation of diversity of human H chain CDR3 regions. *J. Immunol.* **147**, 1720–1729.

Sanz, I., Wang, S.-S., Meneses, G. & Fischbach, M. (1994). Molecular characterization of human Ig heavy chain DIR genes. *J. Immunol.* **152**, 3958–3969.

Shin, E. K., Matsuda, F., Nagaoka, H., Fukita, Y., Imai, T., Yokoyama, K., Soeda, E. & Honjo, T. (1991). Physical map of the 3′ region of the human immunoglobulin heavy chain locus: clustering of autoantibody-related variable segments in one haplotype. *EMBO J.* **10**, 3641–3645.

Siebenlist, U., Ravetch, J. V., Korsmeyer, S., Waldmann, T. & Leder, P. (1981). Human immunoglobulin D segments encoded in tandem multigenic families. *Nature*, **294**, 631–635.

Sonnhammer, E. L. L. & Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **167**, GC1–10.

Staden, R. (1994). The Staden package. In *Methods in Molecular Biology* (Griffin, A. M. & Griffin, H. G., eds), pp. 9–170, Humana Press Inc., Totawa, USA.

Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. (1992). The repertoire of human germline V_H sequences reveals about fifty groups of V_H segments with different hypervariable loops. *J. Mol. Biol.* **227**, 776–798.

Tomlinson, I. M., Cook, G. P., Carter, N. P., Elaswarapu, R., Smith, S., Walter, G., Buluwela, L., Rabbitts, T. H. & Winter, G. (1994). Human immunoglobulin V_H and D segments on chromosomes 15q11.2 and 16p11.2. *Hum. Mol. Genet.* **3**, 853–860.

Tomlinson, I. M., Walter, G., Jones, P. T., Dear, P. H., Sonnhammer, E. L. L. & Winter, G. (1996). The imprint of somatic hypermutation on the repertoire of human germline V genes. *J. Mol. Biol.* **256**, 813–817.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature*, **302**, 575–581.

Tuaillon, N., Miller, A. B., Tucker, P. W. & Capra, J. D. (1995). Analysis of direct and inverted DJ_H rearrangements in a human Ig heavy chain transgenic minilocus. *J. Immunol.* **154**, 6453–6465.

Whitfield, L. S., Hawkins, T. L., Goodfellow, P. N. & Sulston, J. (1995). 41 kilobases of analyzed sequence from the pseudoautosomal and sex-determining regions of the short arm of the human Y chromomsome. *Genomics*, **27**, 306–311.

Wilson, I. A. & Stanfield, R. L. (1994). Antibody-antigen interactions: new structures and new conformational changes. *Curr. Opin. Struct. Biol.* **4**, 857–867.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T. & Cooper, J., *et al.* (1994). 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature*, **368**, 32–38.

Wu, T. T., Johnson, G. & Kabat, E. A. (1993). Length distribution of CDRH3 in antibodies. *Proteins: Struct. Funct. Genet.* **16**, 1–7.

Yamada, M., Wasserman, R., Reichard, B. A., Shane, S., Caton, A. J. & Rovera, G. (1991). Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood B lymphocytes. *J. Exp. Med.* **173**, 395–407.

Zietkiewicz, E., Richer, C., Makalowski, W., Jurka, J. & Labuda, D. (1994). A young Alu subfamily amplified independently in human and African great apes lineages. *Nucl. Acids Res.* **22**, 5608–5612.

Zong, S. Q., Nakai, S., Matsuda, F., Lee, K. H. & Honjo, T. (1988). Human immunoglobulin D segments: isolation of a new D segment and polymorphic deletion of the D_1 segment. *Immunol Letters*, **17**, 329–334.

Edited by J. Karn

(Received 6 March 1997; received in revised form 25 April 1997; accepted 9 May 1997)

Attachment F

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 3576–3580, April 1992
Biochemistry

In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library

(naive immunoglobulin repertoire/filamentous phage/random mutagenesis)

HERMANN GRAM*, LORI-ANNE MARCONI, CARLOS F. BARBAS III, THOMAS A. COLLET, RICHARD A. LERNER, AND ANGRAY S. KANG

The Scripps Research Institute, Departments of Molecular Biology and Chemistry, 10666 North Torrey Pines Road, La Jolla, CA 92037

Contributed by Richard A. Lerner, January 7, 1992

ABSTRACT **We have used a combinatorial immunoglobulin library approach to obtain monoclonal antibodies from nonimmune adult mice, thereby establishing the principles of (*i*) accessing naive combinatorial antibody libraries for predetermined specificities and (*ii*) increasing the affinity of the selected antibody binding sites by random mutagenesis. A combinatorial Fab library expressing immunoglobulin μ and κ light-chain fragments on the surface of filamentous phage was prepared from bone marrow of nonimmunized, adult BALB/c mice with the multivalent display vector pComb8. Phage displaying low affinity Fabs (binding constants, 10^4–10^5 M^{-1}) binding to a progesterone–bovine serum albumin conjugate were isolated from the library. Random mutagenesis of the heavy- and light-chain variable regions expressed in the monovalent phage display vector pComb3 was performed by error-prone PCR, and subsequently clones with improved affinity for the hapten conjugate were selected. We demonstrate that antibodies with desirable characteristics from a nonimmune source may be selected and affinity maturation may be achieved by using the twin vectors pComb8 and pComb3, thus opening the route to obtaining specific antibodies from a generic library and bypassing immunization.**

The B-cell immune response to an antigen can be viewed to occur in two stages. The initial stage generates low-affinity antibodies mostly of the IgM isotype from an existing pool of the B-cell repertoire available at the time of immunization. The second stage, which is driven by antigen stimulation, produces high-affinity antibodies predominantly of the IgG isotype, starting with the light-chain and heavy-chain variable region (V_L and V_H) genes selected in the primary response. The predominant mechanism for affinity maturation is hypermutation of V region genes followed by selection of those cells that produce antibodies of the highest affinity (reviewed in refs. 1 and 2). Both stages in the immune response have been accessed and extensively studied by hybridoma technology (3).

Recently, there has been much interest in using gene cloning approaches to generate and express antibodies by combinatorial library techniques to bypass hybridoma technology (4–19). The main driving force for this approach is to more efficiently harness the vast antibody repertoire. We and others have successfully generated diverse high-affinity antibodies to haptens, virus particles, and protein antigens, thereby recapitulating functional molecules appearing during the natural immune response in animals and humans. Most of the molecular genetic approaches for obtaining active antibodies by screening combinatorial immunoglobulin libraries were directed toward the second stage of the antibody response, since the immunization protocol, the choice of the H-chain IgG isotype for molecular cloning, and/or the high abundance of specific mRNA in the starting material (12) most likely biased the combinatorial libraries toward affinity-matured antibodies. Recently, Marks *et al.* (19) demonstrated that active single-chain (Sc) antibodies with affinity constants between 10^6 and 10^7 M^{-1} against a hapten or a small number of epitopes on a protein can be obtained directly from nonimmune combinatorial immunoglobulin libraries.

In its simplest form, the initial stage of the immune response can be re-created *in vitro* by generating a combinatorial library of PCR-amplified immunoglobulin μ and L chains from the bone marrow of adult mice. This is a close approximation to the naive, unselected repertoire, since the majority of B cells in bone marrow expressing immunoglobulin μ chains have not been subjected to tolerance and antigen selection and should therefore represent all the combinatorial junctional diversity of immunoglobulin V regions (20). The phagemid pComb8 facilitates the display of multiple copies of the Fab fragments along the phage surface permitting access to low-affinity antibodies (15, 16). Hence, specific V_H and V_L pairs could possibly be enriched from a diverse naive repertoire.

The second stage of the immune response *in vivo* involves affinity maturation of the selected specificities by mutation and selection. An efficient way to generate random mutations in the laboratory is by an error-prone replication mechanism, either by targeting the mutations to the antibody binding sites by error-prone PCR (21), or by passaging the phagemid carrying the genetic information for the Fab fragment through an *Escherichia coli mutD* strain, in which the spontaneous mutation frequency is 10^3–10^5 times higher than in a wild-type strain (22). Here we describe our results using error-prone PCR. In subsequent communications, we will describe a complementary method using randomized complementarity-determining region (CDR) 3 sequences of a V_H region to generate mutants (C.F.B., unpublished data). In the present study, we chose progesterone as a hapten since it elicits a restricted immune response in mice (23, 24). The method described here has three essential features: (*i*) the ability to initially access low-affinity Fabs from a naive library by the use of multivalent phage expression systems, (*ii*) subsequent affinity maturation by error-prone PCR, and (*iii*) use of a Sc construct during the maturation process to avoid a high background of artifactual binding due to loss of the L chain. When used in concert, these methods allow selection and affinity maturation of antibodies from a naive library.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: BSA, bovine serum albumin; CDR, complementarity-determining region; gVIII and gIII, M13 gene VIII and gene III, respectively; Sc, single chain; V_L, light-chain variable region; V_H, heavy-chain variable region.

*Present address: Sandoz Pharma, Preclinical Research, CH-4002 Basel, Switzerland.

METHODS

RNA Isolation and cDNA Synthesis. Three nonimmunized adult male (6 months old) BALB/cByJ mice (Scripps breeding colony) were used to prepare 5×10^7 bone marrow cells in 4% fetal calf serum in phosphate-buffered saline (PBS). To deplete for surface IgG-positive cells, the preparation was incubated with rat anti-mouse IgG2b (0.1 ml), goat anti-mouse IgG (0.1 ml), and rabbit anti-mouse IgG2b (0.1 ml) for 30 min at ambient temperature. The cells were pelleted, washed with PBS, and resuspended in PBS (9 ml). Rabbit complement was added (1 ml) and incubated at 37°C for 30 min. The cells were pelleted and total RNA was isolated as described by Kang *et al.* (25). Total RNA was used as a template for cDNA synthesis for μ and κ chains with the following primers: μ chain, 5'-ATTGGGACT AGT TTCTGC GAC AGC TGG ATT-3'; κ chain, 5'-GCG CCG TCT AGA ATT AAC ACT CAT TCC TGT TGAA-3', respectively, using a SuperScript kit (BRL). Briefly, 7 μg of total RNA was mixed with 60 pmol of primer, heated to 70°C for 10 min, and immediately cooled on ice. Two microliters of RNase inhibitor, 10 μl of 5× synthesis buffer, 8 μl of dNTP mixture (to give a final concentration of 200 μM each NTP), 5 μl of 0.1 M dithiothreitol, and 1 μl of BRL SuperScript reverse transcriptase (200 units/μl) were added, and the reaction mixture was made up to 50 μl with diethylpyrocarbonate-treated water. The reaction was allowed to proceed at room temperature for 10 min and then at 42°C for 50 min. The reaction was terminated by incubating at 90°C for 5 min and then by placing on ice for 10 min followed by adding 1 μl of RNase H and incubating at 37°C for 20 min. PCR amplification was performed in a 100-μl reaction mixture as described, using V_H1–9 and the μ-chain primer for the H chains and V_L3–7 and the κ-chain primers for the L chains (25).

Naive Immunoglobulin μ/κ Library Construction. The PCR amplified μ-chain and κ-chain DNAs were cleaved as described (25). The resulting μ-chain *Xho* I/*Spe* I fragments were inserted into the pComb8 phagemid vector (15) to generate a μ-chain–gene VIII (gVIII) fusion library. Transformation into *E. coli* XL1 blue and phage production was carried out essentially as described (26). Subsequently, the κ-chain *Sac* I/*Xba* I fragments were cloned into the H-chain F_d μ–gVIII fusion library.

Panning. The panning procedure used is a modification of that originally described by Parmley and Smith (27). Wells of a microtitration plate (Costar 3690) were coated at 4°C with 50 μl of progesterone-3-(*O*-carboxymethyl)oxime–bovine serum albumin (BSA) conjugate (100 μg/ml) (Sigma P4778) in PBS. The wells were washed twice with water and blocked by completely filling with 1% (wt/vol) BSA in PBS and incubating the plates at 37°C for 1 h. Blocking solution was flicked out and 50 μl of the phage library (typically 10^{11} colony-forming units) in PBS/BSA (0.1%) was added to each well and the plates were incubated for 2 h at 37°C. The washing steps, elution, and multiplication of the phage were done essentially as described (16).

Colony Screening of Panned Libraries. The bacterial colonies were prepared essentially as described (16) and probed for progesterone binding with a progesterone-3-(*O*-carboxymethyl)oxime–horseradish peroxidase conjugate (Sigma P3659). The filters were developed with 4-chloronaphthol (28).

Sc F_V–gIII Fusion. The plasmid pComb3 (16) was digested with the restriction endonucleases *Xba* I and *Nhe* I and religated, thus eliminating the L-chain cloning cassette. A synthetic DNA linker, which encodes a 15-amino acid linker sequence (29) consisting of two oligonucleotides (5'-TCGAGAAAGTCTCTAGAGGTAAATCTTCTGGTTCT-GGTTCCGAATCTAAATCTACTGAGCTCAAAGTCA-3' and 5'-CTAGTGACTTTGAGCTCAGTAGAT-TTAGATCGGAACCAGAACCAGAAGA TTT ACC TCTAGAGACTTTC-3'), was inserted into the *Xho* I/*Spe* I digested truncated pComb3 vector forming the phagemid ScpComb3. The internal recognition sequences for restriction endonucleases *Xba* I (TCTAGA) and *Sac* I (GAGCTC) are underlined. The V_H and V_L segments of the progesterone binders were amplified by PCR using the upstream primer described above and the oligonucleotides 5'-ATTTGG-GAAGGACTGTCTAGATGMRGAGAC-3' (M is A or C and R is A or G) and 5'-GAGGACTAGTTACAGTTGGTGCAG-CATCAG-3', respectively. The internal recognition sequences for *Xba* I (TCTAGA) and *Spe* I (ACTAGT) are underlined. The V_H and V_L F_V PCR fragments were digested with *Xho* I/*Xba* I and *Sac* I/*Spe* I, respectively, and subsequently inserted into ScpComb3.

Expression and Detection of Soluble Fabs and Sc F_V Antibodies. For Fab production, the gVIII moiety in the phagemids encoding the progesterone binders PgA11, PgB6, and PgF1 was excised with restriction endonucleases *Spe* I and *Eco*RI and was subsequently replaced by a synthetic linker encoding a TAA stop codon (underlined). The linker was formed by the oligonucleotides 5'-CTAGTTAACTGAG-TAAG-3' and 5'-AATTCTTACTCAGTTAA-3'. The production and detection of antibody Fab fragment was performed essentially as described (18, 26) except that the *E. coli* cells were disrupted by three freeze–thaw cycles. For producing soluble ScF_v antibody fragments, the V_H–linker–V_L fusions were excised from the ScpComb3 phagemid with *Xho* I and *Spe* I and subcloned into the expression vector pTAC01 (to be described in detail elsewhere), a derivative of pFl260 (30). It has the inducible tac promoter and the *pelB* leader sequence for secretion, and it allows in-frame fusion of the inserted ScF_v with a decapeptide sequence (YPYDVPDYAS) (31) as a tag for immunochemical detection. Expression and detection of the ScF_v antibody fragments was as described above, except that an anti-decapeptide antibody conjugated to alkaline phosphatase was used for the ELISA.

V_H/V_L Targeted Mutagenesis by PCR. Equal amounts of undigested PgF1, PgB6, and PgA11 ScpComb3 plasmids were mixed and serially diluted. Then 100, 10, 1, 0.1, and 0.01 ng of the mixture were subjected separately to 35 cycles (1 min at 94°C, 2 min at 50°C, 1 min at 72°C) of amplification under the following reaction conditions: 50 mM KCl/10 mM Tris·HCl, pH 9.0/6.5 mM $MgCl_2$/0.5 mM $MnCl_2$/0.01% gelatin/0.1% Triton X-100/1 mM each dCTP, dGTP, dTTP/0.2 mM dATP/0.1 mM dITP, using the M13 reverse sequencing primer 5'-AACAGCTATGACCATG-3' and a backward primer complementary to the gIII moiety 5'-GACAGGAG-GTTGAGGCAGGT-3' at 100 μM. The basic method of error-prone PCR was as originally described by Leung *et al.* (21). The PCRs of all template dilutions were pooled and treated with phenol before digestion with *Xho* I and *Spe* I. The gel-purified and digested PCR fragments were ligated back into the *Xho* I/*Spe* I-digested ScpComb3 plasmid. The ligation products were electroporated into *E. coli* XL1 blue giving rise to $\approx 10^6$ transformants. Subsequent steps of phage production and panning were carried out as described above, except that the phage were panned in the absence of BSA.

Affinity Determination. The binding constants of the soluble antibody fragments were determined by competitive ELISA (18, 32). Briefly, wells of a microtitration plate (Costar 3690) were coated at 4°C with 50 μl of progesterone-3-(*O*-carboxymethyl)oxime–BSA conjugate (100 μg/ml) (Sigma P4778) in PBS. The wells were washed twice with water and blocked with 1% BSA in PBS at 37°C for 1 h. Fab or ScF_v supernatants were mixed with progesterone-3-(*O*-carboxymethyl)oxime–BSA in PBS/BSA (0.1%) and incubated in the wells at 37°C for 2 h. The plates were washed with PBS/Tween (0.05%; vol/vol), and goat anti-mouse κ-chain alkaline phosphatase conjugate (Southern Biotechnology Asso-

ciates, Birmingham, AL) or mouse anti-decapeptide monoclonal antibodies conjugated to alkaline phosphatase was added and incubated (37°C; 1 h). The plates were washed as described, and substrate was added [0.1 ml of *p*-nitrophenyl phosphate at 1 mg/ml in 0.1 M Tris·HCl (pH 9.4) containing 50 mM $MgCl_2$]. After incubation (25°C; 60–180 min) the absorbance was read at 405 nm. Apparent affinities were determined as the reciprocal of the hapten concentration required to inhibit 50% maximal binding in a competitive ELISA. This is a close approximation to the affinity (32) and permitted the ranking of the binding activities.

Nucleic Acid Sequencing. The complete nucleotide sequences of the V regions of the H and L chains were determined from double-stranded DNA by using Sequenase 2.0 (United States Biochemical).

RESULTS

Progesterone Binding Fabs from a Naive Murine Repertoire. Bone marrow depleted for surface IgG-positive cells from unprimed mice was chosen as the source for the native antibody repertoire. Total RNA was extracted, and the cDNA fragments encoding the immunoglobulin μ F_d and κ chains were amplified by PCR. A combinatorial library of 5×10^6 members was established by subsequently cloning the μ-immunoglobulin F_d and κ-chain fragments into the vector pComb8, which allows fusion of the H-chain F_d fragment to gVIII. Since the Fab antibody fragments are displayed at a high copy number on the phage surface (15), this vector seemed most suitable for accessing low-affinity antibodies, which are expected to be found in an unselected and unprimed repertoire (2, 33). The recombinant phagemids were packaged into M13 phage particles and five rounds of panning on progesterone-3-(*O*-carboxymethyl)oxime–BSA-coated ELISA wells were performed. Phage eluted after the first and third round were analyzed for expression of anti-progesterone Fabs by bacterial colony lifts. Care was taken to exclude artifacts caused by phagemids expressing the immunoglobulin μ-chain F_d–gVIII fusion without a corresponding L chain. These H-chain-only phages reacted nonspecifically to unrelated antigens (BSA, horseradish peroxidase, hen egg lysozyme), presumably due to the hydrophobic patch displayed on an unpaired H chain (9).

Those colonies producing the strongest signal in the Western blot were further examined, and three clones—PgA11, PgB6, and PgF1—were isolated for subsequent analysis. The first two emerged from the first round of panning, and the latter was isolated after the third round of selection. All three Fabs, produced in their soluble form, bound specifically to progesterone-3-(*O*-carboxymethyl)oxime–BSA and progesterone-11α-hemisuccinyl–BSA. In addition, all three Fabs displayed a significant crossreactivity against an epitope on cytochrome *c* (Fig. 1). Their apparent binding constants for progesterone-3-(*O*-carboxymethyl)oxime–BSA were determined as $\approx 10^4$ M^{-1} for PgA11 and $\approx 3 \times 10^4$ M^{-1} and $\approx 10^5$ M^{-1} for PgF1 and PgB6, respectively. These binding constants are much below that reported for anti-progesterone monoclonal antibodies with affinities of $2\text{–}5 \times 10^8$ M^{-1}. These high-affinity antibodies show little diversity in V gene usage, with an exclusive restriction to the rarely found V_H IX family and the widely used V_{k1} family (23, 24). Since we established and screened the same naive library with the pComb3 vector, previously used to isolate rare high-affinity Fabs (18), without success for progesterone binders (H.G., unpublished results), we reasoned that the cited V_H IX or V_{k1} family or a proper combination of both was not represented in the library. Indeed, none of our progesterone binders utilized V genes of either the V_H IX or V_{k1} family. Our clones PgB6 and PgF1 utilize the same combination of closely related V_H and V_L genes. Both of their V_H genes are identical to two closely related but distinct germ-line genes (34, 35) with no evidence of somatic mutation, as one would expect for a naive repertoire (Fig. 2). The true germ-line gene(s) for their closely related V_L genes are not yet known. Since both the V_H genes have joined to different diversity and joining region segments (see Fig. 2), the clones PgB6 and PgF1 cannot be of the same origin but must have been selected from two independent cloning events, pointing toward a possible importance of this particular combination for progesterone binding. The V_L and V_H genes used by PgA11 are not closely related to the other two clones.

Thus, this part of our study demonstrated that by using a multivalent display vector, Fabs can be isolated from naive combinatorial libraries with affinities comparable to those observed for the primary immune response to haptens, such as phosphocholine (33) and nitrophenol (39). Furthermore, the combinatorial library approach can yield V genes or V gene combinations that would not have been selected *in vivo*.

Affinity Maturation by PCR-Directed Mutagenesis. To mimic the process of somatic mutation, which leads to the selection of antibodies with a higher affinity, we set out to create random mutations in both the V_L and V_H regions and subsequently to select antibodies with an increased affinity to the hapten. To target the mutations by error-prone PCR

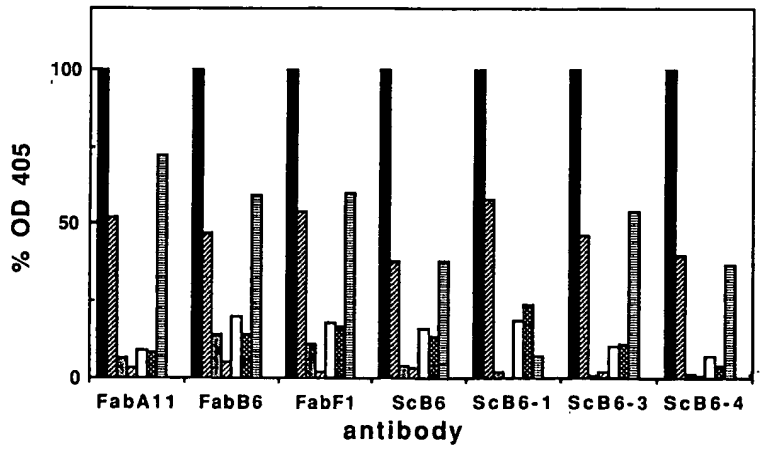


FIG. 1. Crossreactivity of Fabs and Sc antibodies. The reactivity to various antigens was determined by ELISA. Antigens were (from left to right) progesterone-3-(*O*-carboxymethyl)oxime–BSA, progesterone-11α-hemisuccinyl–BSA, horseradish peroxidase, BSA, tetanus toxoid, trypsin inhibitor, and cytochrome c.

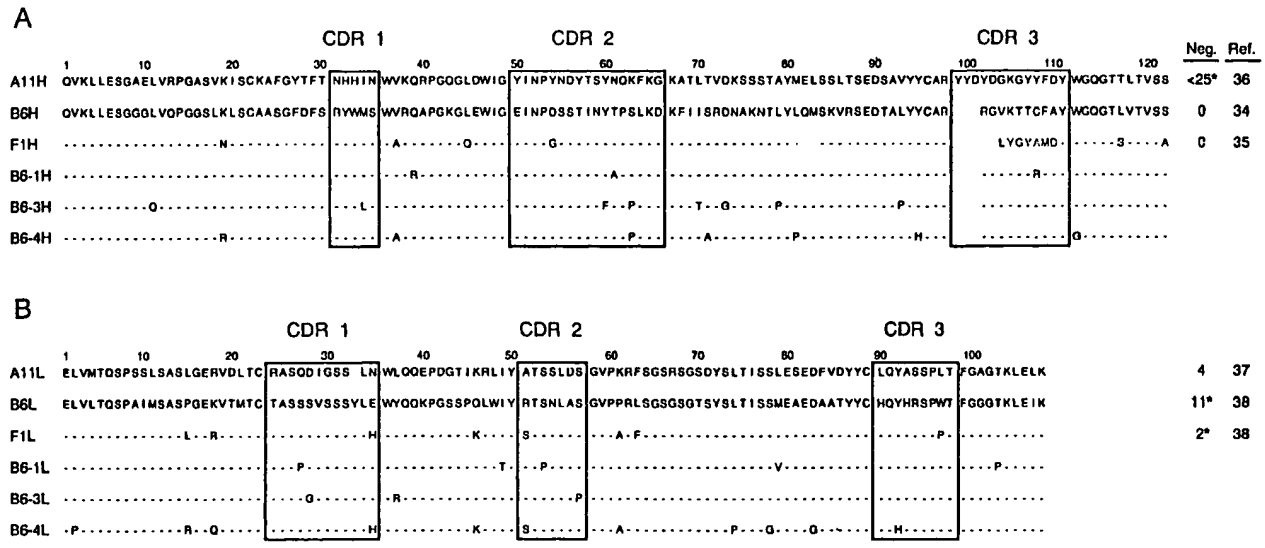


FIG. 2. Amino acid sequences of H-chain (*A*) and L-chain (*B*) of the progesterone binders. CDR sequences are boxed. The number of nucleotide exchanges to the germ-line (neg) gene is given as well as the reference for the most homologous sequence by searching the GenBank 68.0 data base. *, True germ-line gene unknown, closest homology referenced.

specifically and only to the V regions, we created a Sc Fv–gIII fusion plasmid that contained in-frame fusions of the following elements: the *pelB* leader sequence for secretion, the V_H and V_L reading frames linked with a synthetic oligonucleotide coding for a flexible peptide (29), and the gIII moiety. The use of a Sc vector further overcomes the difficulties due to unwanted selection for nonspecific binding by phage that express immunoglobulin H chains only (9).

We chose the monovalent coat protein III display system for the expression and selection of the mutated Sc antibodies, since it facilitates the enrichment for antibodies with higher affinities (16). For mutagenesis, a mixture of the three ScF_V templates was subjected to error-prone PCR amplification. To produce DNA fragments by PCR with various numbers of mutations, various amounts of template DNA were used for the PCRs. The PCR products from all the reactions were pooled before recloning them into the vector ScpComb3. A library of mutated anti-progesterone Sc phage displaying antibody combining sites was established and panned on progesterone-3-(*O*-carboxymethyl)oxime–BSA and the number of eluted colony-forming units was taken as a measure for the relative affinity of the displayed Sc antibodies to the hapten. After the third round of panning, a 50- to 100-fold increase in the yield of eluted phagemids relative to the nonmutated population was noted. Individual mutants showed a 10- to 300-fold increase in yield after panning as compared to the parent clones, indicating that the mutants encoded antibody binding sites with an increased affinity. The four best mutants were chosen for sequence analysis and determination of their affinity for the hapten conjugate.

DNA sequencing revealed that all four mutant clones had arisen from ScPgB6, with ScPgB6-1 and ScPgB6-2 being identical. The predominant type of mutation obtained by this PCR protocol was an A→G/T→C nucleotide exchange (68%), while T→G/A→C, G→A/C→T, T→A/A→T, G→C/C→G, or C→A/G→T exchanges occurred at approximately the same frequency. DNA sequences with a higher than average mutation frequency—i.e., mutational hot spots—were observed. We further noted that the three mutant clones differed in the number of base-pair changes. The mutation frequencies for both the V_H and V_L regions were found to be 1.5% for ScPgB6-1, 2.1% for ScPgB6-3, and 4.1% for ScPgB6-4, which led to multiple amino acid substitutions in the CDRs and framework regions of the mutants (Fig. 2).

The affinity of the mutated Sc antibodies to progesterone-3-(*O*-carboxymethyl)oxime–BSA as determined by competitive ELISA had increased over the parent ScPgB6 antibody by 30-fold for ScPgB6-1 and ≈13-fold for both ScPgB6-3 and ScPgB6-4 (Fig. 3). Interestingly, the clone with the least mutations exhibited the highest affinity. The crossreactivity pattern for the mutant Sc antibodies did not change, except that ScPgB6-1 had lost most of its reactivity to cytochrome *c* (Fig. 1). In extensively studied immune responses to haptens (3, 33, 39), an increase in affinity by 1 order of magnitude could be assigned to specific single amino acid substitutions, which implies that only one or two amino acid exchanges in the combining site of the Sc anti-progesterone antibodies may have accounted for their increased affinity to the hapten conjugate. Since a mutant with only a single amino acid exchange was not recovered, we could not identify the critical residue(s) that caused the enhanced affinity to the hapten conjugate. Furthermore, we did not observe amino acid substitutions common to all three of the mutants, which

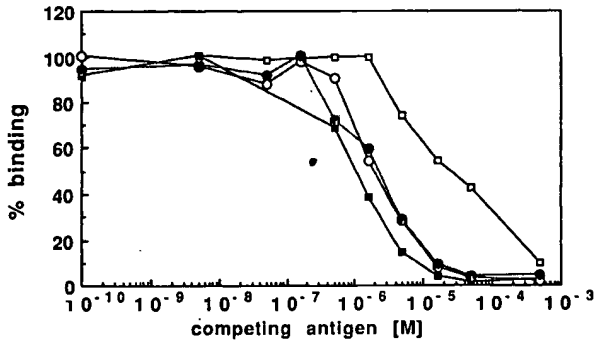


FIG. 3. Relative affinity of Sc anti-progesterone antibodies. The relative affinities of PgScB6 (open box), and the mutants PgScB6-1 (solid box), PgScB6-3 (open circle), and PgScB6-4 (shaded circle) were determined by inhibition ELISA using progesterone-3-(*O*-carboxymethyl)oxime–BSA as competing antigen. Molarity refers to the progesterone moiety on the conjugate, as determined by the supplier.

suggested that changing different residues can account for the increased affinity to 3-(*O*-carboxymethyl)progesterone. A Ser-63 to Pro-63 substitution in the V_H CDR2 is common to the mutants ScPgB6-3 and ScPgB6-4. Although both mutants show a similar affinity to the hapten conjugate, we cannot yet unequivocally assess the importance of that residue for antigen binding, since multiple amino acid exchanges occurred in the V_L and V_H regions of the two mutants.

DISCUSSION

In this study, we present the principle of selection and affinity maturation of specific antibodies to a hapten from a naive library. Although our *in vitro* selection and mutagenesis system is quite simple compared to the complexity of the immune system, we note some features in common. (*i*) The affinity of antibodies selected from a naive combinatorial library can reach the same order of magnitude as antibodies from a primary immune response to haptens. (*ii*) Although the mechanisms generating mutations *in vivo* (40) or *in vitro* were different, we observed mutational hot spots. (*iii*) The increase in affinity after one round of mutation and selection *in vitro* is in the same order of magnitude as that observed for the transition from primary to secondary immune responses to haptens. (*iv*) We could recover a mutated antibody combining site with an altered crossreactivity, as is occasionally observed *in vivo* (39, 41). It will be interesting to explore whether additional cycles of random mutagenesis and selection could lead to a further increase in affinity, as observed, for example, in the tertiary immune response to phenyloxazalone (3).

When the combinatorial antibody approach was first described, one wondered whether it could be used to efficiently tap into the vast antibody repertoire *in vivo* (10). While a number of subsequent reports (11, 12, 14, 18) demonstrated the effectiveness of this approach, the present study shows that we can access and evolve antibody chain combinations that may never be selected *in vivo*. Thus, it now seems as if it is possible to exceed the diversity of the antibody response *in vivo* by molecular cloning techniques.

The ability to recapitulate some features of a primary and secondary immune response to a hapten *in vitro* opens the road to selection of catalytic antibodies that cannot be readily obtained from an immune source and suggests that circumventing immunization for generation of high-affinity antibodies may soon be a reality.

We would like to thank Dr. Norman Klinman and Dr. Robert Chanock for critical reading of the manuscript and Dr. Ian Wilson for helpful discussion. We are also grateful to Dr. Ruth Epstein for technical advice and to Terry Jones for assistance.

1. Griffith, G. M., Berek, C., Kaartinen, M. & Milstein, C. (1984) *Nature (London)* **312**, 271–275.
2. Kock, C. & Rajewsky, K. (1989) *Annu. Rev. Immunol.* **7**, 537–559.
3. Berek, C. & Milstein, C. (1987) *Immunol. Rev.* **96**, 23–41.
4. Skerra, A. & Pluckthun, A. (1988) *Science* **240**, 1038–1041.
5. Better, M., Chang, C. P., Robinson, R. R. & Horowitz, A. H. (1988) *Science* **240**, 1041–1043.
6. Orlandi, R., Gussow, D. H., Jones, P. T. & Winter, G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 3833–3837.
7. Chiang, Y. L., Dong, R., Brow, M. A. & Larrick, J. W. (1989) *Biotechniques* **7**, 360–366.
8. Sastry, L., Alting-Mees, M., Huse, W. D., Short, J. M., Sorge, J. A., Hay, B. N., Janda, K. D., Benkovic, S. J. & Lerner, R. A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5728–5732.
9. Ward, E. S., Gussow, D., Griffiths, A. D., Jones, P. T. & Winter, G. (1989) *Nature (London)* **341**, 544–546.
10. Huse, W. D., Sastry, L., Iverson, S., Kang, A. S., Alting-Mees, M., Burton, D. R., Benkovic, S. J. & Lerner, R. A. (1989) *Science* **246**, 1275–1281.
11. Caton, A. J. & Koprowski, H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6450–6454.
12. Persson, M. A., Cao, R. H. & Burton, D. R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2432–2436.
13. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990) *Nature (London)* **348**, 552–554.
14. Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1991) *Nature (London)* **352**, 624–628.
15. Kang, A. S., Barbas, C. F., Janda, K. D., Benkovic, S. J. & Lerner, R. A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4363–4366.
16. Barbas, C. F., Kang, A. S., Lerner, R. A. & Benkovic, S. J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7978–7982.
17. Kang, A. S., Jones, T. M. & Burton, D. R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 11120–11123.
18. Burton, D. R., Barbas, C. F., Persson, M. A., Koenig, S., Chanock, R. M. & Lerner, R. A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10134–10137.
19. Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D. & Winter, G. (1991) *J. Mol. Biol.* **222**, 581–597.
20. Decker, D. J., Boyle, N. E., Koziol, J. A. & Klinman, N. R. (1991) *J. Immunol.* **146**, 350–361.
21. Leung, D. W., Chen, E. & Goeddel, D. V. (1989) *J. Methods Cell Mol. Biol.* **1**, 11–15.
22. Fowler, R. G., Schaaper, R. M. & Glickman, B. W. (1986) *J. Bacteriol.* **167**, 130–137.
23. Deverson, E., Berek, C., Taussig, M. & Feinstein, A. (1987) *Eur. J. Immunol.* **17**, 9–13.
24. Stura, E. A., Arevalo, J. H., Feinstein, A., Heap, R. B., Taussig, M. J. & Wilson, I. A. (1987) *Immunology* **62**, 511–521.
25. Kang, A. S., Burton, D. R. & Lerner, R. A. (1991) *Methods: Comp. Methods Enzymol.* **2**, 111–118.
26. Barbas, C. F. & Lerner, R. A. (1991) *Methods: Comp. Methods Enzymol.* **2**, 119–124.
27. Parmley, S. F. & Smith, G. P. (1988) *Gene* **73**, 305–318.
28. Nakane, P. K. (1968) *J. Histochem. Cytochem.* **16**, 557–560.
29. Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S. M., Pope, S. H., Riordan, G. S. & Whitlow, M. (1988) *Science* **242**, 423–426.
30. Larimer, F. W., Mural, R. J. & Soper, T. S. (1990) *Prot. Eng.* **3**, 227–231.
31. Field, J., Nikawa, J. I., Broek, D., MacDonald, B., Rodgers, L., Wilson, I. A., Lerner, R. A. & Wigler, M. (1988) *Mol. Cell. Biol.* **8**, 2159–2165.
32. Rath, S., Stanley, C. M. & Steward, M. W. (1988) *J. Immunol. Methods* **106**, 245–249.
33. Malipiero, U., Levy, N. S. & Gearhart, P. J. (1987) *Immunol. Rev.* **96**, 59–75.
34. Ollo, R., Auffray, C., Sikorav, J. L. & Rougeon, F. (1981) *Nucleic Acids Res.* **9**, 4099–4109.
35. Hartmann, A. B. & Rudikoff, S. (1984) *EMBO J.* **3**, 3023–3030.
36. Klein, D., Nietupski, J., Sirlin, S. & Stavnezer, J. (1988) *J. Immunol.* **140**, 1676–1684.
37. Seidmann, J. G., Max, E. E. & Leder, P. (1979) *Nature (London)* **280**, 370–375.
38. Ewulonu, U. K., Nell, L. J. & Thomas, J. W. (1990) *J. Immunol.* **144**, 3091–3098.
39. Allen, D., Cumano, A., Dildrop, R., Kocks, C., Rajewsky, K., Rajewsky, N., Roes, J., Sablitzki, F. & Siekevitz, M. (1987) *Immunol. Rev.* **96**, 6–22.
40. Steele, E. J., Pollard, J. W., Taylor, L. & Both, G. W. (1991) in *Somatic Hypermutation in V-Regions*, ed. Steele, E. J. (CRC, Boca Raton, FL), pp. 137–149.
41. Diamond, B. & Scharff, M. D. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5841–5844.

Attachment G

Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 3809–3813, April 1994
Medical Sciences

In vitro evolution of a neutralizing human antibody to human immunodeficiency virus type 1 to enhance affinity and broaden strain cross-reactivity

(passive immunization/viral neutralization/phage display/combinatorial libraries/complementarity-determining regions)

CARLOS F. BARBAS III*†, DANA HU*, NANCY DUNLOP‡, LYNETTE SAWYER§, DOUG CABABA*, R. MICHAEL HENDRY§, PETER L. NARA‡, AND DENNIS R. BURTON*¶

Departments of *Molecular Biology and ¶Immunology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037; ‡Laboratory of Tumor Cell Biology, Virus Biology Section, National Cancer Institute–Frederick Cancer Research and Development Center, Frederick, MD 21702; and §Viral and Rickettsial Disease Laboratory, California Department of Health Services, 2151 Berkeley Way, Berkeley, CA 94704

Communicated by Robert M. Chanock, January 4, 1994

ABSTRACT **A method is described that allows for the improvement of antibody affinity. This method, termed complementary-determining region (CDR) walking, does not require structural information on either antibody or antigen. Complementary-determining regions are targeted for random mutagenesis followed by selection for fitness, in this case increased binding affinity, by the phage-display approach. The current study targets a human CD4-binding-site anti-gp120 antibody that is potently and broadly neutralizing. Evolution of affinity of this antibody demonstrates in this case that affinity can be increased while reactivity to variants of human immunodeficiency virus type 1 is broadened. The neutralizing ability of this antibody is improved, as assayed with laboratory and primary clinical isolates of human immunodeficiency virus type 1. The ability to produce human antibodies of exceptional affinity and broad neutralizing ability has implications for the therapeutic and prophylactic application of antibodies for human immunodeficiency virus type 1 infection.**

The ability to clone human antibodies in large numbers from seropositive individuals (1, 2) or to create them *de novo* by using synthetic approaches (3–5) promises increased application of this class of molecules in the service of human health. There are a number of considerations in choosing an antibody-combining site; these are primarily affinity and specificity. Current molecular methods should allow for experimenter-controlled evolution of binding sites to satisfy demands in both areas. The generation of molecules with exceptional affinities should both increase biological potency and decrease the cost of antibodies as therapeutics.

Currently, there is an increased urgency for the development of molecules for the prophylaxis and therapy of human immunodeficiency virus type 1 (HIV-1) infection. Passive immunotherapy has been successfully used against a number of viruses (6) and indeed has been used to protect chimpanzees against HIV-1 infection (7, 8) and to protect cynomolgus monkeys against simian immunodeficiency virus and HIV type 2 infection (9). One of the major problems in using antibodies as anti-HIV-1 reagents is sequence variation in the envelope proteins of the virus. Because the virus requires the binding of the surface glycoprotein gp120 to the CD4 molecule on the target cell for infectivity (10, 11), the CD4-binding site on gp120 has become a popular target for antiviral antibodies (12–16). However, antibodies to this region are not generally particularly potent in terms of virus neutralization. Furthermore, such antibodies tend to be even less potent against primary isolates of virus than the more commonly used laboratory-adapted strains (17). Using combinatorial libraries, we have isolated human anti-CD4-binding site antibodies with quite exceptional neutralizing ability (18). Nevertheless, we wished to improve the likelihood that these antibodies could succeed in prophylactic and therapeutic application.

In the present report we develop a strategy for evolution of antibody affinity. The method is applied to an HIV-1 neutralizing human antibody directed against the CD4-binding site of gp120. For this antibody, which already shows exceptional neutralizing potency, we show the possibility of increasing affinity, potency, and broadening strain reactivity.

MATERIALS AND METHODS

Reagents, Strains, and Vectors. Oligonucleotides were from Operon Technologies (Alameda, CA). *Escherichia coli*, phage, and the phagemid vector pComb3 are as described (19). The recombinant glycoproteins (rgps) 120 IIIB(LAI) and 120 MN were purchased from American Bio-Technologies (Cambridge, MA) and AgMed (Cambridge, MA), respectively. Reagents for surface plasmon resonance experiments were obtained from Pharmacia.

Library Construction and Selection. *Experiment A.* A *Hin*dIII restriction site was introduced preceding the heavy-chain complementarity-determining region (CDR) I by standard methods into clone HIV-4 (18) in the pComb3 vector (19). Clone HIV-4 (Fab b4) and HIV-12 (Fab b12) are identical. This Fab was selected by panning against rgp120 IIIB(LAI). The GenBank accession no. of HIV-4 sequence is L03147. A CDRI library was constructed by PCR of the above construct with primers (*i*) 5′-GAA-GGT-TTC-TTG-TCA-AGC-TTC-TGG-ATA-CAG-ATT-CAG-TNN-SNN-SNN-SNN-SNN-STG-GGT-GCG-CCA-GGC-CCC-C and (*ii*) primer R3B(20), where N is A,C,G, or T, and S is G or C.

The PCR product was gel-purified, digested with *Hin*dIII and *Spe* I, and gel-purified. The product was ligated with *Hin*dIII- and, *Spe* I-digested HIV-4. Subsequent steps were as described (3, 17) to produce phage displaying antibody Fab fragments on their surface. The library, 2×10^7 clones, was affinity-selected by four rounds of panning against gp120 IIIB(LAI) immobilized on Costar 3690 microtiter wells. One well coated with 1 μg of gp120 IIIB was used for each round of selection. After selection, plasmid DNA was prepared, and individual clones were sequenced.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CDR, complementarity-determining region; HCDR1 and HCDR3, heavy chain CDR1 or CDR3, respectively; HIV-1, human immunodeficiency virus type 1; PBMC, peripheral blood mononuclear cell; sCD4, soluble CD4; rgp, recombinant glycoprotein.

†To whom reprint requests should be addressed.

Experiment B. Plasmid DNA isolated after experiment A was used as a template for PCR with oligonucleotide primers (*iii*) CCC-TTT-GCC-CCA-GAC-GTC-CAT-ATA-ATA-ATT-GTC-CTG-GGG-AGA-ATC-ATC-MNN-MNN-MNN-MNN-CCC-CAC-TCT-CGC-ACA and (*iv*) FTX-3 (20) to randomize within heavy chain CDR3 (HCDR3) (M = A or C, which gives K = T or G in the complementary strand; NNK doping strategy). The PCR product was gel-purified, digested with *Aat* II and *Xho* I, and gel-purified. This PCR product was ligated with *Aat* II, *Xho* I-digested HIV-4. Subsequent steps were described above to yield a library of 8×10^6 clones. The library was affinity-selected by six rounds of panning against gp120 IIIB. After selection, the gIII fragment was removed, and soluble Fab was produced. The complete amino acid sequences of the variable regions of selected antibodies were deduced by dideoxynucleotide chain-termination sequencing. Fab was purified to homogeneity by affinity chromatography, as described (19).

Surface Plasmon Resonance. The kinetic constants for the binding of Fab to rgp120 IIIB and MN were determined by surface plasmon resonance-based measurements using the BIAcore instrument from Pharmacia. The sensor chip was activated for immobilization with *N*-hydroxysuccinimide and *N*-ethyl-*N'*-(3-diethylaminopropyl)carbodiimide. The proteins, rgp120 IIIB or MN, were coupled to the surface by injection of 50 μl of a 50 μg/ml sample. Excess activated esters were quenched with 15 μl of ethanolamine (1 M and pH 8.5). Typically 4000 resonance units were immobilized. Binding of Fab fragments to immobilized gp120 was studied by injection of Fab in a range of concentrations (0.5–10 μg/ml) at a flow rate of 5 μl/min. The association was monitored as the increase in resonance units per unit time. Dissociation measurements were acquired after the end of the association phase but with a flow rate of 50 μl/min. The binding surface was regenerated with HCl (1 M NaCl and pH 3) and remained active for 20–40 measurements. The association and dissociation rate constants, k_{on} and k_{off}, were determined from a series of measurements as described (20–22). Equilibrium association and dissociation constants were deduced from the rate constants.

Quantitative Infectivity Assay Based on Syncytium Formation. Quantitative neutralization assays with the MN and LA1 (IIIB) strains were done as described (23). Monolayers of CEM-SS target cells were cultured with virus in the presence or absence of Fab, and the number of syncytium-forming units of input virus was determined 3–5 days later. Equivalent amounts of virus were used in the assays to allow direct comparison of Fab concentrations tested. Data represent the average of at least two runs. Assays were repeatable over a virus-surviving fraction range of 1–0.001 within a 2- to 4-fold difference in the concentration of antibody ($P < 0.001$).

Microplaque Neutralization. The quantitative measurement of the reduction of infectivity of primary clinical isolates of HIV-1 was determined with a microplaque assay, as described (24). MT2 cells were used as indicator cells in this assay. The isolation of HIV-1 from frozen peripheral blood lymphocytes obtained from seropositive donors has been described (25). A number of the primary isolates of HIV-1 used in this study have been described (26); these are VL135, VL434, VL069, VL263, and VL596, previously described as isolates 1, 3, 4, 5, and 7, respectively.

RESULTS

CDR Walking. In experiment A, the entire heavy chain CDR (HCDR1), as defined by Kabat *et al.* (27), was targeted for mutagenesis using the overlap PCR mutagenesis strategy described (3). NNS- or NNK-type doping strategies were used with no assumptions made as to the most fit residue at each position. After four rounds of selection for binding to rgp120 IIIB, the sequencing of 12 clones indicated a preference for asparagine (N) at position 31, an aromatic residue at position 32, serine (S) or threonine (T) primarily at position 33, branched hydrophobic residues at position 34, and hydrophobic and/or aromatic residues at position 35 (Fig. 1). Experiment B introduced diversity into HCDR3 at positions 96–99 of the clones that survived the four rounds of selection of experiment A. After six rounds of selection for binding rgp120 IIIB, a strong consensus was seen in both mutagenized CDRs (Fig. 1). At the time these selection experiments were done, only rgp120 IIIB was commercially available. Only in case 3B8 was the starting HCDR3 nucleotide and amino acid sequence identical to the parent, indicating some contamination in the secondary library. The net results of this two-step sequential CDR walk were minimal changes from the starting clone. The parental residues at positions 31, 32, 34, and 99 were strongly or absolutely maintained. The hydrophobic parental residue Val-33 was predominantly hydrophilic threonine or serine after selection. Position 96 appears flexible to a variety of substitutions as does position 98. Position 97 shows a preference for the increased steric bulk of the Tyr-97 → Trp mutation.

Affinity Measurement. After selections, four clones were chosen for further study. Clones were chosen that were related to one another by small changes in amino acid sequence and which displayed the most dramatic change in amino acid identity at positions 96 and 98. The kinetics of binding of purified Fab to two types of rgp120 from the highly divergent isolates MN and IIIB were chosen (28). Comparison of the protein rgp120 MN to rgp120 IIIB revealed 88 amino acid changes in the aligned sequences from rgp120 IIIB, as well as 11 deletions and 5 insertions of amino acids. Binding kinetics were studied in real time by using surface plasmon resonance. The kinetic and calculated equilibrium constants are tabulated for the binding of Fabs to both rgp120s (Table 1). The parental clone HIV-4 binds rgp120 IIIB with ≈10-fold better affinity than rgp120 MN, a trend that is maintained for the evolved clones. The highest-affinity Fab, 3B3, is improved 8-fold in affinity to rgp120 IIIB and 6-fold in affinity to rgp120 MN. The increases in affinities of the evolved Fabs binding to rgp120 IIIB and rgp120 MN are well-correlated as shown in Fig. 2. Thus, without selective pressure for binding to rgp120 MN, increase in affinity to

Experiment A	Experiment B		
CDR1	CDR1	CDR3	
31 32 33 34 35	31 32 33 34 35	96 97 98 99	
N F V I H		P Y S W	HIV-4
R Y T V F	N F T L M	Q W N W	3B1
N W S V M	N Y T I M	P W T W	3B2
G Y T L M	N F T V H	E W G W	3B3
N F T L L	N Y T L I	P W N W	3B4
H Y S L M	N F I I M	L W N W	3B6
N W V V H	N F S I M	S W R W	3B7
N F S I M	N Y T I Q	P Y S W	3B8
N F A I H	N F T V H	P W R W	3B9
N F T M V			
N F T L Q			
Y F T M H			
S Y P L H			

FIG. 1. CDR walking for the selection of improved variants of HIV-1. In experiment A, HCDR1, is randomized over positions 31–35. After selection for binding to rgp120 IIIB the sequences listed in experiment A were observed. Experiment B introduces additional diversity into HCDR3 positions 96–99 in clones that were selected in experiment A. After additional selective pressure to bind rgp120 IIIB, the sequences listed under experiment B were observed. The sequences of the parental clone HIV-4 are shown for comparison.

Table 1. Binding and neutralization data for evolved Fab reacting with laboratory isolates of HIV-1

Fab	gp120 type	k_{on}, $M^{-1}s^{-1}$	k_{off}, s^{-1}	K_a, M^{-1}	K_d, M	IC_{50} μg/ml	IC_{50} M
HIV-4	IIIB	7.6×10^4	4.8×10^{-4}	1.6×10^8	6.3×10^{-9}	3.9×10^{-2}	7.7×10^{-10}
HIV-4	MN	3.4×10^4	1.5×10^{-3}	2.3×10^7	4.4×10^{-8}	3.0×10^{-1}	5.9×10^{-9}
3B1	IIIB	8.5×10^4	1.1×10^{-4}	7.7×10^8	1.3×10^{-9}	2.2×10^{-2}	4.4×10^{-10}
3B1	MN	1.4×10^5	1.8×10^{-3}	7.8×10^7	1.3×10^{-8}	9.2×10^{-3}	1.9×10^{-10}
3B3	IIIB	8.4×10^4	6.5×10^{-5}	1.3×10^9	7.7×10^{-10}	4.7×10^{-2}	9.4×10^{-10}
3B3	MN	1.6×10^5	1.2×10^{-3}	1.3×10^8	7.5×10^{-9}	5.5×10^{-3}	1.1×10^{-10}
3B4	IIIB	7.7×10^4	3.6×10^{-4}	2.1×10^8	4.8×10^{-9}	5.0×10^{-2}	9.9×10^{-10}
3B4	MN	8.6×10^4	4.1×10^{-3}	2.1×10^7	4.8×10^{-8}	2.0×10^{-2}	3.9×10^{-10}
3B9	IIIB	4.5×10^4	1.8×10^{-4}	2.5×10^8	5.0×10^{-9}	6.6×10^{-2}	1.3×10^{-9}
3B9	MN	8.1×10^4	1.1×10^{-3}	7.4×10^7	1.4×10^{-8}	7.8×10^{-3}	1.6×10^{-10}

The ability of parental and evolved Fabs to bind rgp120 IIIB and MN was determined by surface plasmon resonance (20–22). The equilibrium association and dissociation constants were calculated from the experimentally determined kinetic constants where $K_a = k_{on}/k_{off}$ and $K_d = k_{off}/k_{on}$. The interpolated IC_{50} values of Fab-neutralizing MN and IIIB viral stocks, as determined with the quantitative infectivity assay based on syncytium formation (23), are given in μg/ml and in molar units.

rgp120 IIIB is accompanied by increased affinity to rgp120 MN. Though a series of single-point mutations is not available in the clones examined to assign changes in affinity directly, comparison of Fab 3B3 with Fab 3B9 and Fab 3B1 with Fab 3B4, which each differ at two positions, suggests change of Pro-96 to glutamine or glutamate as the most productive change. Further examination of acidity changes within the evolved Fabs suggests affinity increases are correlated with decreased pI values of Fabs. The antigens rgp120 IIIB and MN have basic calculated pI values of 9.5 and 9.3, respectively. pI considerations may also contribute to the anomalous behavior of Fab 3B9, which is the only Fab that is increased in pI, as compared with the parent.

Neutralization Studies. Quantitative neutralization assays with the laboratory-adapted strains MN and LAI (IIIB) were done to determine potency of the Fabs (23). As shown in Fig. 3*A*, the evolved Fabs are clearly improved in their abilities to neutralize infectivity of the MN viral stock. The binding affinity of Fabs to rgp120 MN is well-correlated with ability to neutralize the MN stock (Fig. 4). The highest-affinity Fab, 3B3, is improved 54-fold with respect to neutralization of the MN isolate in this assay (Table 1). A different MN viral stock was used in these studies than had been used in the initial characterization of Fab HIV-4 (18). This result is reflected in a difference in activity, as compared with this previous report. Studies with the LAI (IIIB) viral stock show a clustering of Fabs with similar potencies (Fig. 3*B*). With this viral stock a range of reactivity of only 3-fold is noted with the most potent Fab, 3B1, showing a modest 2-fold increase in potency. With both viral stocks, Fabs demonstrate exceptional potency in the 10^{-10} M range (Table 1). To further

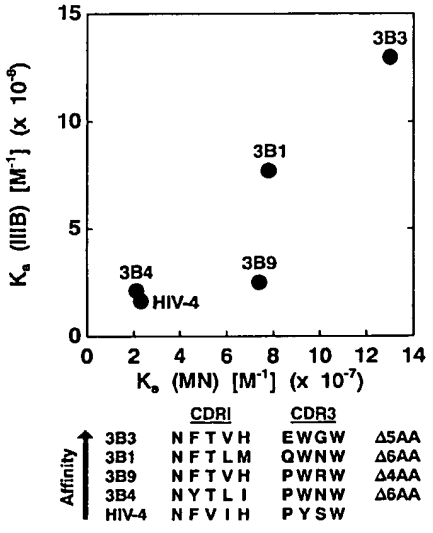


FIG. 2. The affinity increases of evolving Fabs for binding the divergent envelope proteins rgp120 IIIB and rgp120 MN are well-correlated. Affinities were determined by using the surface plasmon resonance technique (20–22). The sequences of evolved clones are ranked as compared with the parent, and changes in the amino acid (AA) sequence from the parent are shown as ΔAA. See also Table 1.

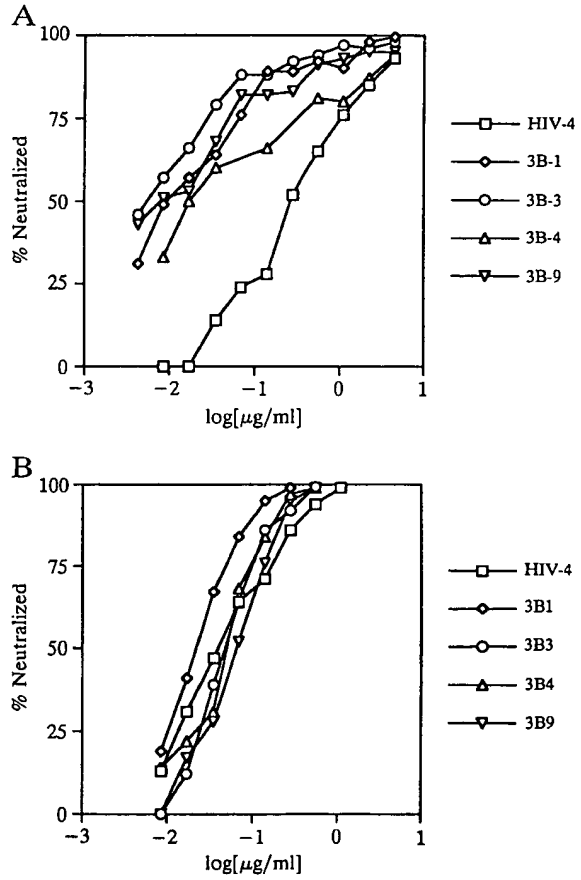


FIG. 3. Evolved Fabs are potent in neutralization assays with laboratory isolates of HIV-I. (*A*) Parental and improved Fabs are compared in a quantitative infectivity assay based on syncytium formation (23) with MN viral stock. (*B*) Comparison with LAI (IIIB) viral stock. Results indicate the average of at least two assays.

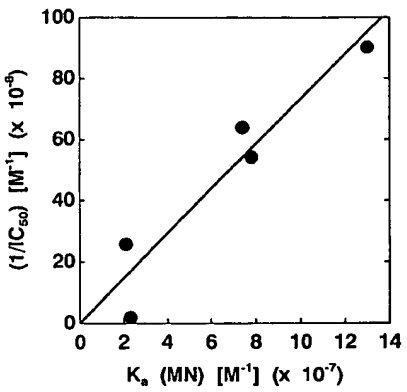


FIG. 4. Affinity increases are well correlated with increased potency in the quantitative infectivity assay with the MN viral stock. IC_{50} values for the reduction of infectivity were interpolated from Fig. 3*A*. The equilibrium association constant for binding rgp120 MN was determined as in Table 1.

delineate functional changes that accompany increased affinity, neutralization studies with primary clinical isolates of HIV-1 were done. Primary clinical isolates of HIV-1 were grown in peripheral blood mononuclear cells (PBMCs) (25). As controls in this microplaque assay (24), IIIB and MN stocks, as well as isolate VL069, were propagated in H9 cells. Additionally, assays with all stocks were done with a pooled plasma from 13 HIV-1-seropositive individuals. In all cases MT2 cells were used as indicator cells. The results are shown in Table 2. In these assays, the highest-affinity Fab, 3B3, is now able to neutralize an additional four isolates, as compared with the parent Fab, HIV-4. Fifty-percent neutralization of isolates VL135 and VL530 by Fab 3B3 at 38.9 and 29.5 μg/ml, respectively, is significant because the parent, Fab HIV-4, showed insignificant levels of neutralization, ≈10%, at 50 μg/ml. Neutralization of these isolates with pooled positive plasma shows that these isolates are relatively resistant to neutralization, as compared with the laboratory isolates grown in H9 cells. Neutralization of MN and IIIB stocks by antibody 3B3 is improved ≈5-fold in this assay. A host-cell effect is noted with isolate VL069. Propagation of this isolate in H9 cells results in a sensitization to neutralization. This effect has been noted previously and is discussed in detail elsewhere (29). In these assays the intrinsic error of the interpolated titers averages ±30%.

Table 2. Neutralization of primary clinical isolates of HIV-1 with natural and evolved Fab

Virus	Host cell	Amount for 50% neutralization, μg/ml		Titer
		HIV-4	3B3	+ PHP
VL135	PBMC	>50	38.9	1:33
VL263	PBMC	17.0	6.6	<1:10
VL596	PBMC	33.1	17.0	1:10
VL069	PBMC	>50	>50	<1:10
VL434	PBMC	>50	10.5	1:10
VL114	PBMC	>50	5.2	<1:10
VL172	PBMC	>50	>50	1:10
VL530	PBMC	>50	29.5	<1:10
VL750	PBMC	>50	>50	1:10
IIIB	H9	0.36	0.068	1:767
MN	H9	0.18	0.044	1:24,000
VL069	H9	3.6	3.5	1:1,200

The potency of the HIV-4 parental clone and Fab 3B3 to neutralize primary clinical isolates was measured in a microplaque assay (24). Virus was either propagated in PBMC or H9 cells. The neutralizing ability of pooled human plasma from 13 HIV-I-seropositive patients (+PHP) is shown for comparison as the titer of the serum dilution. Furthermore, the laboratory-adapted stocks IIIB and MN were also tested in this variant assay for comparison. A host-cell effect is shown for isolate VL069 grown in H9 cells (29).

DISCUSSION

The present study shows the feasibility of improving antibody affinity and function where specific structural information on both antibody and antigen is not available and the antibody already possesses high affinity. The current approach termed "CDR walking" is a variant of our synthetic-antibody approach for the generation of additional specificities *in vitro* (3–5). Practically, there is one important difference. CDR walking involves a limited introduction of diversity into the CDR regions of a defined antibody, as contrasted with the synthetic approach for the generation of new specificities where structural diversity is stressed over library completeness. Diversity in the present case is limited to 6 or fewer amino acid residues with an NNK- or NNS-doping strategy so as to ensure near-complete representation of all possible amino acid combinations. Selection from the library with the phage-display technique then allows for the refinement of the contact between antigen and antibody, which may result from unpredictable sequence changes in the region of interest. Repeated introduction of diversity into CDRs followed by stringent selections should allow for the refinement of human antibodies to levels of affinity far beyond those generated by the immune response. Two strategies are evident for the application of this approach, either sequential or parallel optimization of CDRs. Parallel optimizations makes the assumption that the optimized loops will exhibit additivity in free energy changes when the individually optimized loops are combined (30, 31). In many cases, additivity will likely be observed. Sequential optimization takes into account that additivity may not always be observed and that optimal binding may result from the interdependence of loops. Such interdependence could result from coordinated structural changes on binding antigen and is supported by recent evidence that suggests induced-fit mechanisms may best describe antibody–antigen recognition (32, 33). The two-step sequential walk reported here demonstrates the potential of this strategy. Both sequential and parallel approaches are being examined at present. In this initial study HCDR1 and HCDR3 were chosen for optimization. These CDRs were targeted because rearrangements of these CDRs have been observed on binding of another antibody to antigen (33). Residues 96–99 of HCDR3 were targeted because chain-shuffling experiments indicated this region is a hotspot during the natural maturation of this antibody (34). After the randomization and selection protocol that sampled mutations of 9 amino acid positions, higher-affinity Fabs resulted that had mostly modest changes from the parent (Fig. 1). The most radical change involves the Pro-96 → Glu mutation found in the highest-affinity clone 3B3, which is 8-fold improved in binding rgp120 IIIB. Interestingly, this mutation is also observed in the *in vivo* response, as revealed by chain shuffling. The *in vivo* antibodies that share this mutation are, however, of lower affinity than HIV-4 and have accumulated many additional somatic mutations throughout their sequence (34).

The key issue in producing antibodies to HIV-1 for therapeutic or prophylactic purposes is that they should be highly potent (of high affinity and neutralizing ability) and be cross-reactive with a wide range of isolates. These are usually two opposing characteristics. We have chosen HIV-4, as it recognizes a conformational epitope on gp120 that overlaps the CD4-binding site of gp120 and is broadly and potently active (19). If HIV-4 is truly recognizing the conserved features (shape) of the CD4-binding region, it should be

possible to increase its affinity to gp120 for many or all viral strains, as to date all HIV-1 isolates use CD4 as their primary receptor. This phenomenon is, indeed, observed as shown in Fig. 2. Binding has been increased to MN and IIIB, two highly divergent isolates (28). Selective pressure could have been applied to favor cross-reactivity by selecting with a mixture of divergent gp120s; however, this did not prove to be necessary in the present investigation. The present strategy was dictated because only rgp120 IIIB was commercially available when the selection experiments were done. Potency as judged by quantitative neutralization assays with MN and LAI (IIIB) stocks is improved as well (Fig. 3). With the MN isolate affinity is well-correlated with neutralizing ability (Fig. 4). Neutralization of MN and LAI stocks with soluble CD4 (sCD4) revealed IC_{50} values of 0.6 nM and 0.8 nM, respectively (35). As shown in Table 1, the parental clone and Fab 3B1 have IC_{50} values of 0.8 nM and 0.4 nM, respectively, for the LAI stock. For MN, the parental clone and Fab 3B3 have IC_{50} values of 6 nM and 0.1 nM, respectively. The ability of these evolved monovalent Fabs to neutralize with potencies equivalent to sCD4 is distinctive. The lack of correlation of rgp120 IIIB affinity with neutralizing ability may reflect the sensitivity of the assay conditions in this range. In a recent multicenter study of human and mouse anti-HIV-1 antibodies, no bivalent antibody has demonstrated such potency (36).

Can a single CD4-site antibody fulfill the promises once made by sCD4 as a therapeutic agent? Primary clinical isolates often require 1000-fold more sCD4 for neutralization than laboratory isolates (37–39). This fact may be the primary contributor to the failure of sCD4. As shown in Table 2, the highest-affinity Fab also demonstrates improved ability to neutralize primary clinical isolates. Four isolates not neutralized by the parent are now neutralized by Fab 3B3. For isolate VL114 a titration with bivalent CD4 IgG, predicted to have high activity, yielded an IC_{50} of 10 μg/ml, as compared with 5.2 μg/ml for Fab 3B3. As most of the CDR residues have yet to be optimized, it should be possible to further evolve this Fab to affinities 100 or 1000 times those reported here. These preliminary results suggest that broadly reactive antibodies of exceptional affinity can be prepared. Such antibodies will likely be of use at least to prevent vertical transmission of virus and in cases of accidental exposure. Ideally a mixture of such antibodies directed against several epitopes would be used. It remains to be demonstrated whether antibodies alone can be effective in cases where lymph nodes and thymus are seeded with the virus; however, they will likely be valuable components in combination therapies (40) and perhaps as key targeting agents in future therapies.

We thank Richard A. Lerner for support and encouragement of this project; Terri Jones and Kim Green for protein purification; and Pamela Morden, Mary Kate Morris, and Leo Ocequerq for technical support in primary isolate neutralization assays. C.F.B. is a Scholar of the American Foundation for AIDS Research and the recipient of a Scholar award from the Cancer Research Institute. Support for this work was provided by National Institutes of Health Grant AI33292 and Johnson & Johnson.

1. Williamson, R. A., Burioni, R., Sanna, P.-P., Partridge, L. J., Barbas, C. F., III, & Burton, D. R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4141–4145.
2. Burton, D. R. & Barbas, C. F., III (1993) in *Protein Engineering of Antibody Molecules for Prophylactic and Therapeutic Applications in Man*, ed. Clark, M. (Academic Titles, Nottingham, U.K.), pp. 65–82.
3. Barbas, C. F., III, Bain, J. D., Hoekstra, D. M. & Lerner, R. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4457–4461.
4. Lerner, R. A., Kang, A. S., Bain, J. D., Burton, D. R. & Barbas, C. F., III (1992) *Science* **258**, 1313–1314.
5. Barbas, C. F., III, Languino, L. R. & Smith, J. W. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10003–10007.
6. Burton, D. R. (1991) in *Vaccines '93*, eds. Ginsberg, H. S., Brown, F., Chanock, R. M. & Lerner, R. A. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 1–5.
7. Emini, E. A., Schleif, W. A., Nunberg, J. H., Conley, A. J., Eda, Y., Tokoyoshi, S., Putney, S. D., Matsushita, S., Cobb, K. E., Jett, C. M., Eichberg, J. W. & Murthy, K. K. (1992) *Nature (London)* **355**, 728–730.
8. Ward, R. H., Capon, D. J., Jett, C. M., Murthy, K. K., Mordenti, J., Lucas, C., Frie, S. W., Prince, A. M., Green, J. D. & Eichberg, J. W. (1991) *Nature (London)* **352**, 434–436.
9. Putkonen, P., Thorstensson, R., Ghavamzadeh, L., Albert, J., Hild, K., Biberfeld, G. & Norrby, E. (1991) *Nature (London)* **352**, 436–438.
10. Dalgleish, A. G., Beverley, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F. & Weiss, R. A. (1984) *Nature (London)* **312**, 763–767.
11. Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J. C. & Montagnier, L. (1984) *Nature (London)* **312**, 767–768.
12. Sun, N.-C., Ho, D. D., Sun, C. R. Y., Liou, R.-S., Gordon, W., Fung, M. S. C., Li, X.-L., Ting, R. C., Lee, T.-H., Chang, N. T. & Chang, T. W. (1989) *J. Virol.* **63**, 3579–3585.
13. Thali, M., Olshevsku, U., Furman, C., Gabuzda, D., Posner, M. & Sodroski, J. (1991) *J. Virol.* **65**, 6188–6193.
14. Tilley, S. A., Honnen, W. J., Racho, M. E., Hilgartner, M. & Pinter, A. (1991) *Res. Virol.* **142**, 247–259.
15. Karwowska, S., Gorny, M. K., Buchbinder, V., Giankakos, C., Williams, T. F. & Zolla-Pazner, S. (1992) *AIDS Res. Hum. Retroviruses* **8**, 1099–1106.
16. Moore, J. P. & Ho, D. D. (1993) *J. Virol.* **67**, 863–875.
17. Moore, J. P. & Sweet, R. W. (1993) *Perspectives in Drug Discovery and Design* **1**, 235–250.
18. Barbas, C. F., III, Bjorling, E., Chiodi, F., Dunlop, N., Cababa, D., Jones, T. M., Zebedee, S. L., Persson, M. A. A., Nara, P. L., Norrby, E. & Burton, D. R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9339–9343.
19. Barbas, C. F., III, Kang, A. S., Lerner, R. A. & Benkovic, S. J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7978–7982.
20. Barbas, C. F., III, Amberg, W., Simoncsits, A., Jones, T. M. & Lerner, R. A. (1993) *Gene* **137**, 57–62.
21. Altschuh, D., Dubs, M.-C., Weiss, E., Zeder-Lutz, G. & Van Regenmortel, M. H. V. (1992) *Biochemistry* **31**, 6298–6304.
22. Karlsson, R., Michaelsson, A. & Mattsson, L. (1991) *J. Immunol. Methods* **145**, 229–240.
23. Nara, P. L., Hatch, W. C., Dunlop, N. M., Robery, W. G., Arthur, L. O., Gonda, M. A. & Fischinger, P. J. (1987) *AIDS Res. Hum. Retroviruses* **3**, 283–302.
24. Hanson, C. V., Crawford-Miksza, L. & Sheppard, H. W. (1990) *J. Clin. Microbiol.* **28**, 2030–2034.
25. Gallo, D., Kimpton, J. S. & Dailey, P. J. (1987) *J. Clin. Microbiol.* **25**, 1291–1294.
26. Wrin, T., Crawford, L., Sawyer, L., Weber, P., Sheppard, H. W. & Hanson, C. V. (1994) *J. Acquired Immune Defic. Syndr.* **7**, 211–219.
27. Kabat, E. A., Wu, T. T., Perry, H. & Gottesman, K. (1991) *Sequences of Proteins of Immunological Interest* (U.S. Dept. of Health and Human Services, Washington, DC), 5th Ed.
28. Myers, G., Korber, B., Berzofsky, J. A., Smith, R. F. & Pavlakis, G. N. (1992) *Human Retroviruses and AIDS 1992* (Theoretical Biology and Biophysics, Los Alamos, NM).
29. Sawyer, L. S. W., Wrin, M. T., Crawford-Miksza, L., Potts, B., Wu, Y., Weber, P. A., Alfonso, R. D. & Hanson, C. V. (1994) *J. Virol.* **68**, 1342–1349.
30. Wells, J. A. (1990) *Biochemistry* **29**, 8509–8517.
31. Lowman, H. B., Bass, S. H., Simpson, N. & Wells, J. A. (1991) *Biochemistry* **30**, 10832–10838.
32. Wilson, I. A. & Stanfield, R. L. (1993) *Curr. Opin. Struct. Biol.* **3**, 113–118.
33. Stanfield, R. L., Takimoto-Kamimura, M., Rini, J. M., Profy, A. T. & Wilson, I. A. (1993) *Structure* **1**, 83–93.
34. Barbas, C. F., III, Collet, T. A., Amberg, W., Roben, P., Binley, J. M., Hoekstra, D., Cababa, D., Jones, T. M., Williamson, R. A., Pilkington, G. R., Haigwood, N. L., Cabezas, E., Satterthwait, A. C., Sanz, I. & Burton, D. R. (1993) *J. Mol. Biol.* **230**, 812–823.
35. Layne, S. P., Merges, M. J., Dembo, M., Spouge, J. L. & Nara, P. L. (1990) *Nature (London)* **346**, 277–279.
36. D'Souza, M. P., Geyer, S. J., Hanson, C. V., Hendry, M., Milman, G. & Collaborating Investigators. (1994) *AIDS* **8**, 169–181.
37. Daar, E. S., Li, X. L., Moudgil, T. & Ho, D. D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6574–6578.
38. Moore, J. P., McKeating, J. A., Huang, Y., Ashkenazi, A. & Ho, D. D. (1992) *J. Virol.* **66**, 235–243.
39. Orloff, S. L., Kennedy, M. S., Belperron, A. A., Madon, P. J. & McDougal, J. S. (1993) *J. Virol.* **67**, 1461–1471.
40. Fauci, A. (1993) *Science* **262**, 1011–1018.

Attachment H

J. Mol. Biol. (1996) 263, 551–567

JMB

Isolation of Picomolar Affinity Anti-c-erbB-2 Single-chain Fv by Molecular Evolution of the Complementarity Determining Regions in the Center of the Antibody Binding Site

Robert Schier[1]*, Adrian McCall[2], Gregory P. Adams[2] Keith W. Marshall[1], Hanne Merritt[1], Michael Yim[1], Robert S. Crawford[1] Louis M. Weiner[2], Cara Marks[1] and James D. Marks[1]*

[1]*Departments of Anesthesia and Pharmaceutical Chemistry, University of California, San Francisco SFGH Rm 3C-38, 1001 Potrero Avenue, San Francisco, CA 94110, USA*

[2]*Department of Medical Oncology, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia PA 19111, USA*

**Corresponding author*

We determined the extent to which additional binding energy could be achieved by diversifying the complementarity determining regions (CDRs) located in the center of the antibody combining site of C6.5, a human single-chain Fv (scFv) isolated from a non-immune phage library which binds the tumor antigen c-erbB-2. CDR3 of the light (V_L) and heavy (V_H) chain variable region of C6.5 were sequentially mutated, the mutant scFv displayed on phage, and higher affinity mutants selected on antigen. Mutation of V_L CDR3 yielded a scFv (C6ML3-9) with a 16-fold lower K_d (1.0×10^{-9} M) than C6.5. Due to its length of 20 amino acids, four V_H CDR3 libraries of C6ML3–9 were constructed. The greatest increase in affinity from a single library was ninefold ($K_d = 1.1 \times 10^{-10}$ M). Combination of mutations isolated from separate V_H CDR3 libraries yielded additional ninefold decreases in K_d, resulting in a scFv with a 1230-fold increase in affinity from wild-type C6.5 ($K_d = 1.3 \times 10^{-11}$ M). The increase in affinity, and its absolute value, are comparable to the largest values observed for antibody affinity maturation *in vivo* or *in vitro* and indicate that mutation of V_L and V_H CDR3 may be a particularly efficient means to increase antibody affinity. This result, combined with the location of amino acid conservation and substitution, suggests an overall strategy for *in vitro* antibody affinity maturation. In addition, the affinities and binding kinetics of the single-chain Fv provide reagents with potential tumor targeting abilities not previously available.

© 1996 Academic Press Limited

Keywords: c-erbB-2; single-chain Fv; antibody phage display; affinity maturation; BIAcore

Introduction

Antibody based cancer therapy is limited by the properties of antibodies derived from conventional hybridoma technology (reviewed by Riethmueller *et al.*, 1993). IgG are large molecules (150 kDa) which diffuse slowly into tumors (Clauss & Jain, 1990) and are slowly cleared from the circulation, resulting in poor tumor:normal organ ratios (Sharkey *et al.*, 1990). Smaller single-chain Fv

Present address: R. Schier, Codon Genetic Systems, Nussdorfer Laende 11, 1190 Vienna, Austria.

Abbreviations used: AMP, ampicillin; c-erbB-2 ECD, c-erbB-2 extracellular domain; CDR, complementarity determining region; ELISA, enzyme linked immunosorbent assay; FACS, fluorescence activated cell sorter; FR, framework region; Glc, glucose; HBS, Hepes buffered saline (10 mM Hepes, 150 mM NaCl, pH 7.4); IMAC, immobilized metal affinity chromatography; k_{on}, association rate constant; k_{off}, dissociation rate constant; MPBS, skimmed milk powder in PBS; MTPBS, skimmed milk powder in TPBS; PBS, phosphate buffered saline (25 mM NaH_2PO_4, 125 mM NaCl, pH 7.0); PCR, polymerase chain reaction; RU, resonance units; scFv, single-chain Fv fragment; TPBS, 0.05% (v/v) Tween 20 in PBS; SPR, surface plasmon resonance; V_κ, immunoglobulin kappa light chain variable region; V_λ, immunoglobulin lambda light chain variable region; V_L, immunoglobulin light chain variable region; V_H, immunoglobulin heavy chain variable region; wt, wild-type.

0022-2836/96/440551-17 $25.00/0

© 1996 Academic Press Limited

antibody fragments (scFv, 25 kDa) penetrate tumors better than IgG (Yokota *et al.*, 1992), are cleared more rapidly from the circulation, and provide greater targeting specificity (Adams *et al.*, 1993; Colcher *et al.*, 1988; Milenic *et al.*, 1991). scFv are typically constructed from the heavy (V_H) and light (V_L) chain variable region genes of murine IgG, and thus are still potentially immunogenic. In addition, scFv are monovalent and dissociate from tumor antigen faster than bivalent IgG molecules, which exhibit a higher apparent affinity due to avidity (Crothers & Metzger, 1972). Loss of avidity, combined with rapid clearance from blood, results in significantly lower quantitative retention of scFv in tumor (Adams *et al.*, 1992). Significant tumor retention beyond 24 hours will require a dissociation rate constant (k_{off}) less then $10^{-4}\,s^{-1}$ ($t_{1/2} = 1.8$ hours). Since antibodies typically have rapid ($>10^5\,M^{-1}\,s^{-1}$) association rate constants (k_{on}), this requires a K_d ($<10^{-9}$ M) rarely achievable by murine immunization (Foote & Eisen, 1995).

The limitations of hybridoma technology can be overcome by the display of natural (Marks *et al.*, 1991) and synthetic immunoglobulin variable region gene repertoires (Hoogenboom & Winter, 1992) on the surface of filamentous bacteriophage (Hoogenboom *et al.*, 1991; McCafferty *et al.*, 1990). Human scFv can be recovered from these libraries against virtually any antigen (Griffiths *et al.*, 1993; Marks *et al.*, 1991, 1993; Vaughan *et al.*, 1996). Using this approach, we isolated a human scFv (C6.5) which binds to the extracellular domain (ECD) of the tumor antigen c-erbB-2 (McCartney *et al.*, 1995) with a K_d of 1.6×10^{-8} M and k_{off} of $6.3 \times 10^{-3}\,s^{-1}$ (Schier *et al.*, 1995). Biodistribution studies in *scid* mice demonstrate high tumor:normal organ ratios and excellent tumor visualization, however quantitative delivery of scFv to tumor is inadequate to provide therapeutic dosimetry. Greater delivery should be possible by engineering higher affinity scFv.

Phage display can also be used to increase affinity (Lowman *et al.*, 1991; Marks *et al.*, 1992). The antibody sequence is diversified and higher affinity binders selected from the mutant antibody library (Barbas *et al.*, 1994; Hawkins *et al.*, 1992; Yang *et al.*, 1995). For this work, we demonstrate that restriction of mutagenesis to the complementarity determining regions (CDRs) located in the center of the antibody combining site can yield increases in affinity comparable to values previously reported either for *in vivo* or *in vitro* affinity maturation. Mutation of the V_L and V_H CDR3 of C6.5 scFv yielded a scFv with a 1230-fold increased affinity ($K_d = 1.3 \times 10^{-11}$ M). The decrease in K_d of mutant scFv was largely due to a reduction of k_{off}, which correlated well with the $t_{1/2}$ observed for retention on the surface of c-erbB-2 expressing SK-OV-3 cells. Modeling of the location of mutations suggests a general approach for the rapid and efficient generation of ultra-high affinity antibodies.

Results

Mutation of C6.5 scFv V_L CDR3

Library construction and selection

For construction of a library of C6.5 V_L CDR3 mutants, an oligonucleotide was designed (Table 1) which partially randomized nine amino acid residues located in V_L CDR3 (residues 89 to 95b, numbering according to Kabat *et al.*, 1991; Table 2). For the nine amino acids randomized, the ratio of nucleotides was chosen so that the frequency of wild-type (wt) amino acid was 49%. After transformation, a library of 1.0×10^7 clones was obtained. The mutant phage antibody library was designated C6VLCDR3. Polymerase chain reaction

Table 1. Sequences of primers used

VL1	5'-GTCCCTCCGCCGAACACCCA,5,2,2,5,3,1,6,1,3,5,3,1,7,4,2,7,4,2,2,2,1,5,3,2,5,3,2, ACAGTAATAATCAGCCTCAT-3'
VL2	5'-GAGTCATTCTCGACTTGCGGCCGCACCTAGGACGGTCAGCTTGGTCCCTCCGCCGAACACCCA-3'
VHA	5'-GCGCAGTTGGAACTACTGCA,5,8,8,5,8,8,5,8,8,5,8,8, ATGTCTCGCACAAAAATACACGGC-3'
RVHA	5'-TGCAGTAGTTCCAACTGCGC-3'
VHB	5'-GTATTCAGGCCACTTTGCGCA,5,8,8,5,8,8,5,8,8,5,8,8, GCAATATCCCACGTCATGTC-3'
RVHB	5'-TGCGCAAAGTGGCCTGAATAC-3'
VHC	5'-CTGGCCCCAATGCTGGAAGTA,5,8,8,5,8,8,CCA,5,8,8,5,8,8,GCAGTTGGAACTACTGCAATATCC-3'
RVHC	5'-TACTTCCAGCATTGGGGCCAG-3'
VHD	5'-GACCAGGGTGCCCTGGCCCCA,5,8,8,5,8,8,5,8,8,5,8,8,TTCAGGCCACTTTGCGCAGTTGG-3'
RVHD	5'-TGGGGCCAGGGCACCCTGGTC-3'
C6hisnot	5'-GATACGGCACCGGCGCACCTGCGGCCGCATGGTGATGATGGTGATGTGCGGCACCTAGGACGGTCAGCT TGG-3'
PML3-9	5'-CCTAGGACGGTCAGCTTGGTCCCTCCGCCGAACACCCAACCACTCAGGGTGTAATCCCAGGATGCACAG TAATAATCAGC-3'
PML3-12	5'-CCTAGGACGGTCAGCTTGGTCCCTCCGCCGAACACCCAACCACTCCGGCTGTAATCCCATGCTGCACAG-3'
PCD1	5'-GACGGTGACCAGGGTGCCCTGGCCCCAAACGTGCAGCCATTCAGGCCACTTTGCGCA-3'
PCD2	5'-GACGGTGACCAGGGTGCCCTGGCCCCATACGCCCAGCCATTCAGGCCACTTTGCGCA-3'
PCD3	5'-GACGGTGACCAGGGTGCCCTGGCCCCAGTTGTCCAACCATTCAGGCCACTTTGCGCA-3'
PCD5	5'-GACGGTGACCAGGGTGCCCTGGCCCCACATCTGCATCCATTCAGGCCACTTTGCGCA-3'
PCD6	5'-GACGGTGACCAGGGTGCCCTGGCCCCAGGGGTACATCCATTCAGGCCACTTTGCGCA-3'

Nucleotide mixtures used, molar fraction: 1 A (0.7), C, G, and T (0.1); 2 C (0.7), A, G, and T (0.1); 3 G (0.7); 4 T (0.7), A, C, and G (0.1); 5 C and G (0.5); 6 C (0.7) and G (0.3); 7 C (0.3) and G (0.7); 8 A, C, G, and T (0.25).

(PCR) screening revealed that 30 of 30 randomly selected colonies had full length insert and diversity was confirmed by sequencing the V_L CDR3 of ten unselected clones (results not shown). Prior to selection, 5/92 clones selected at random expressed scFv which bound c-erbB-2 ECD by enzyme linked immunosorbent assay (ELISA). After both the third and fourth rounds of selection, 92/92 clones bound c-erbB-2 ECD by ELISA.

The C6VLCDR3 library was selected using decreasing concentrations of biotinylated c-erbB-2 ECD, as described by Schier *et al.* (1996b). A relatively high antigen concentration (4.0×10^{-8} M) was used for the first round to capture rare or poorly expressed phage antibodies. The antigen concentration was then decreased 40-fold for the second round (1.0×10^{-9} M), and decreased a further tenfold each of the subsequent two rounds (1.0×10^{-10} M, third round; 1.0×10^{-11} M, fourth round). Reduction of the antigen concentration helps ensure that selection for higher affinity scFv occurs, rather than selection for scFv that express well on phage or are less toxic to *Escherichia coli* (Hawkins *et al.*, 1992; Schier *et al.*, 1996b). The optimal antigen concentration cannot be predicted *a priori*, due to variability in phage antibody expression levels and uncertainty regarding the highest affinities present in the mutant phage antibody library. Thus the choice of antigen concentration was guided by determining the percentage of binding phage present in the polyclonal phage preparation. After each round of selection, polyclonal phage were prepared and the concentration of binding phage determined by measuring the rate of binding to c-erbB-2 ECD under mass transport limited conditions using surface plasmon resonance (SPR) in a BIAcore (Schier & Marks, 1996). The percentage of binding phage was calculated by dividing the concentration of binding phage by the concentration of total phage as determined by infecting *E. coli*. If little or no change in the binding phage percentage was observed (as in these experiments), the antigen concentration was decreased significantly (at least tenfold) in the next round of selection. A large decrease in the percentage of binding phage, as we have observed during other selections (Schier & Marks, 1996), indicates that the selection should be repeated using a higher antigen concentration, or that the antigen concentration should not be decreased for the next round of selection.

Characterization of mutant scFv

To identify scFv with a lower K_d than wt scFv, apparent k_{off} was determined by SPR in a BIAcore on unpurified native scFv in bacterial periplasm (Schier *et al.*, 1996b). A total of 24 scFv from the third and fourth rounds of selection were ranked by k_{off}. After the third round of selection, 80% of scFv had a lower k_{off} than wt and after four rounds, 100% of scFv had a lower k_{off} than wt scFv. The 12 scFv with the lowest k_{off} from each of these rounds of selection were sequenced and each unique scFv gene was subcloned for purification. scFv were purified by immobilized metal affinity chromatography (IMAC), followed by gel filtration to remove any dimeric or aggregated scFv. The k_{on} and k_{off} were determined by BIAcore, and the K_d calculated.

After the third round of selection, seven unique scFv were identified, all with higher affinity than wt scFv (Table 2). scFv had on average 1.8 amino acid substitutions/scFv, with a single substitution at residue 93 the most frequently observed mutation. This single amino acid substitution would have occurred with a frequency of 1/12,000 in the original library, assuming equal nucleotide coupling efficiency. The average scFv affinity was 3.6×10^{-9} M (4.4-fold increase), with the highest affinity 2.6×10^{-9} M (six fold increase). After four rounds of selection, six scFv were identified, and none of these sequences was observed in the scFv sequenced from the third round (Table 2). For the selections reported above, binding phage were not specifically eluted, but rather were incubated with *E. coli*. We subsequently determined that when elutions are performed by incubating phage bound to antigen with *E. coli*, the phage probably must dissociate from antigen for infection to occur, leading to preferential selection of scFv with more rapid k_{off} (Schier & Marks, 1996). Steric hindrance, due to the size of paramagnetic beads, blocks the attachment of pIII on antigen bound phage to the f-pilus on *E. coli*. Repetition of the fourth round of selection using 100 mM HCl as eluent yielded an additional five scFv, including the highest affinity scFv obtained (C6ML3-9, Table 2). scFv from the fourth rounds of selection had on average 2.9 amino acid substitutions/scFv, with expected frequencies between 1/590,000 and 1/24,000,000 in the original library. The average scFv affinity after the fourth rounds of selection was 1.9×10^{-9} M (8.4-fold increase), with the highest affinity 1.0×10^{-9} M (16-fold increase). The results demonstrate the efficiency and importance of the selection and elution techniques for isolating very rare high affinity clones from a library.

Location of mutations in higher affinity scFv

Significant sequence variability (six different amino acids) was observed at residues 93 and 94, with less variability (three different amino acids) at residues 95 and 95a. Thus a subset of the randomized residues appear to be more important in modulating affinity. All but one of these four residues (L95) appear to have solvent accessible side-chains in our model of C6.5, which is based on the atomic structure of the Fab KOL (Figure 1). Three of the residues randomized (A89, W91, and G96) were 100% conserved in all mutants sequenced. Two additional residues (A90S and D92E) showed only a single conservative substitution. These conserved residues appear to have a structural role in the variable domain, either in maintaining the main-chain conformation of the

Table 2. Sequences, affinities and binding kinetics of scFv isolated from a library of C6.5 V_L CDR3 mutants

Clone	F	V_L CDR3 sequence	K_d (10^{-9} M)	k_{on} (10^5 s^{-1} M^{-1})	k_{off} (10^{-3} s^{-1})
		8 9 9 9 5ab 7			
C6.5	0	AAWDDSLSGWV	16.0	4.0 ± 0.20	6.3 ± 0.06
A. *Third round of selection*					
C6ML3-7	1	----YAV----	2.6	6.5 ± 0.29	1.7 ± 0.09
C6ML3-2	2	----H------	2.8	7.0 ± 0.24	2.0 ± 0.09
C6ML3-6	2	-S--Y------	3.2	5.8 ± 0.43	1.9 ± 0.02
C6ML3-4	1	-S-EY--W---	3.4	3.8 ± 0.32	1.3 ± 0.13
C6ML3-5	4	----Y------	3.7	5.2 ± 0.34	1.9 ± 0.08
C6ML3-3	1	-S--YR-----	3.8	5.5 ± 0.12	2.1 ± 0.05
C6ML3-1	1	----Y--W---	6.1	3.3 ± 0.07	2.0 ± 0.15
B. *Fourth round of selection*					
C6ML3-9[a]	1	-S--YT-----	1.0	7.6 ± 0.20	0.76 ± 0.03
C6ML3-14[a]	1	-----P-W---	1.1	7.0 ± 0.20	0.77 ± 0.02
C6ML3-23[a]	1	-S--H--W---	1.5	6.7 ± 0.41	1.7 ± 0.09
C6ML3-19[a]	1	-S--RP-W---	1.5	6.6 ± 0.69	1.0 ± 0.02
C6ML3-12[a]	2	----Y-R----	1.6	4.5 ± 0.16	0.72 ± 0.02
C6ML3-29	1	----GT-W---	1.7	12.9 ± 1.03	2.2 ± 0.02
C6ML3-15	1	----RP-W---	2.2	5.9 ± 0.81	1.3 ± 0.02
C6ML3-10	1	---E-P-Y---	2.3	6.1 ± 0.80	1.4 ± 0.02
C6ML3-13	1	----AT-W---	2.4	8.7 ± 0.98	2.1 ± 0.09
C6ML3-8	1	----HLRW---	2.6	6.4 ± 0.23	1.7 ± 0.15
C6ML3-11	1	----YA-W---	3.6	6.1 ± 0.15	2.2 ± 0.08

A, Mutants isolated after the third round of selection; B, mutants isolated after the fourth round of selection. The entire V_L CDR3 of C6.5 is shown, with the residues subjected to mutagenesis (89 to 95b) underlined. k_{on}, and k_{off} were measured by SPR in a BIAcore, and the K_d calculated. Dashes indicate sequence identity. F, Frequency of isolated scFv. Numbering is according to Kabat *et al.* (1987).

[a] scFv obtained after elution with 100 mM HCl or 100 mM triethylamine.

loop, or in packing on the V_H domain. Residues A89, W91, and D92 are identical in both C6.5 and KOL (Marquart *et al.*, 1980), with conservative substitutions A90S and G96A observed at the other two positions in KOL, consistent with a structural role. In the model of C6.5, G95b is in a turn and A89, A90, and W91 are either buried or pack against the V_H domain at the V_H-V_L interface (Figure 1). Hydrogen bonds between V_λD92 and V_λS27a and V_λN27b bridge L3 and L1 to stabilize the L3 and L1 conformations.

Mutation of C6ML3-9 scFv V_H CDR3

Library construction and selection

To further increase the affinity of C6.5, we chose to mutate the V_H CDR3 of the highest affinity scFv (C6ML3-9, $K_d = 1.0 \times 10^{-9}$ M) isolated from the C6VLCDR3 library, rather than mutate C6.5 V_H CDR3 independently and combine mutants. This sequential approach was taken since the kinetic effects of independently isolated antibody fragment mutations are frequently not additive (Yang *et al.*, 1995; Schier *et al.*, 1996b). Due to the length of the C6.5 V_H CDR3 (20 amino acids), a high resolution functional scan was performed on C6.5 scFv in an attempt to reduce the number of amino acids subjected to mutation. Residues 95 to 99, 100a to 100d, and 100g to 102 were separately mutated to alanine, and the K_d of the mutated scFv determined. Residue 100f (alanine) was not studied. Residues 100 and 100e are a pair of cysteines residues separated by four amino acids. A homologous sequence in KOL (Marquardt *et al.*, 1980) results in a disulfide bond between the two cysteine residues and a four residue miniloop. Therefore the two cysteine residues were simultaneously mutated to serine. Results of the alanine scan are shown in Table 3. No detectable binding to c-erbB-2 ECD could be measured by BIAcore for C6.5H95A, C6.5W100hA, and C6.5E100jA. Three additional alanine mutants (G98A, Y100kA, and F100lA) yielded scFv with 20-fold to 100-fold higher K_d than wt scFv. Substitution of the two cysteine residues by alanine (100, 100e) yielded an scFv with a 17.5-fold higher K_d, and a much faster k_{off} (1.38×10^{-1} s^{-1}) than wt C6.5. The remainder of the alanine substitutions yielded only minor (0.5 to 3.7-fold) increases or decreases in K_d.

Based on the results of the alanine scan and a model of C6.5 based on the Fab KOL (Marquardt *et al.*, 1980), residues H95, C100, and C100e were not mutated due to their probability of having an important structural role. H95 is likely to be buried at the V_H-V_L interface where it makes critical packing contacts with the V_L domain. The two cysteine residues also are likely to have a structural role in maintaining the miniloop conformation. W100h was also not mutated given the unique features of tryptophan in antibody combining sites

(Mian *et al.*, 1991). The remaining 16 amino acids were completely randomized four residues at a time in four separate C6VHCDR3 libraries (96 to 99, library A; 100a to 100d, library B; 100f, 100g, 100i, and 100j, library C, and 100k to 102, library D; see Table 4). After transformation, libraries were obtained with sizes 1.7×10^7 (library A), 1.3×10^7 (library B), 3.0×10^6 (library C), and 2.4×10^7

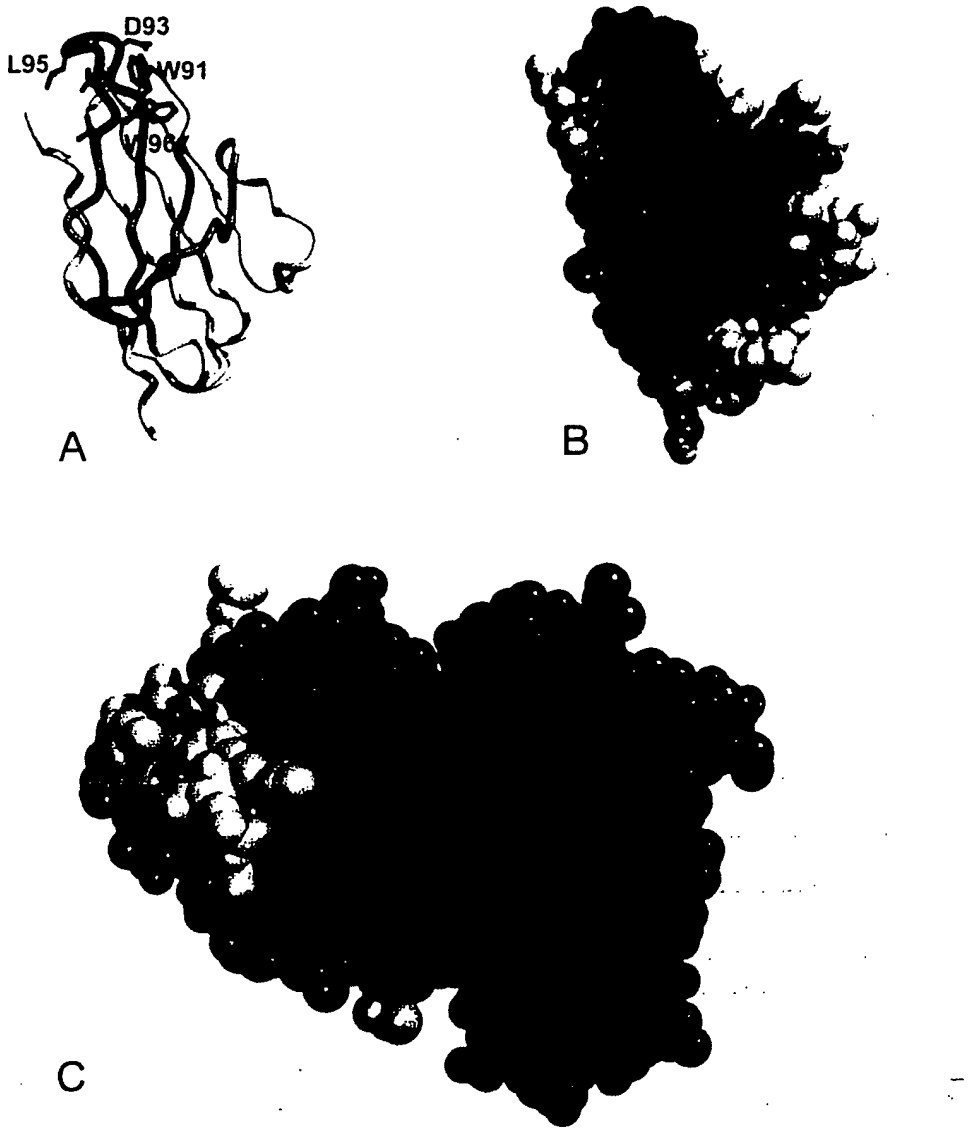


Figure 1. Model of the location of mutations in V_λ CDR3. The location of mutations present in higher affinity scFv isolated from the V_L CDR3 library were modeled on the structure of the Fab fragment of the immunoglobulin KOL (Marquart *et al.*, 1980). A, V_λ domain of KOL, rotated to view the four β-strands (yellow) that pack on the V_H domain at the V_H-V_L interface. V_L CDR3 residues which are conserved in mutant scFv are shown in green (A89, A90, W91, D92, and G 95b) and those that are not conserved and modulate affinity are shown in red (D93, S94, L95, and S95a). CDR3 residues which are contributed by the joining (J) gene segment were not mutated and are shown in blue (W 96 and V 97). Framework residues comprising the four β-strands that pack at the V_H-V_L interface are shown in yellow. Conserved residues extend the β-strands, while non-conserved amino acids form a four-residue loop. B, Same view as A, but with side-chains represented in space filling format. Side-chains of conserved residues (A89, A90, W91) pack directly or indirectly at the V_H-V_L interface. Non-conserved residues have solvent accessible side-chains. C, View from above, looking down on the antibody combining site. The V_H domain is dark blue (right) and the V_L domain is light gray (left). V_H CDR1, V_H CDR2, V_L CDR1, and V_L CDR2 are colored magenta. Conserved V_L CDR3 residues (green) are buried, at the V_H-V_L interface, and by the V_H CDR3 (dark gray) except for a portion of the side-chain of W91. Non-conserved residues (red) have solvent accessible side-chains, with D93, the most frequently mutated residue, located in the center of the binding pocket.

Table 3. Binding kinetics of C6.5 V_H CDR3 mutants obtained by alanine scanning

scFv clone	K_d(mutant)/ K_d (C6.5)	K_d (10^{-9} M)	k_{on} (10^5 M^{-1} s^{-1})	k_{off} (10^{-2} s^{-1})
C6.5H95A	NB	NB	NB	NB
C6.5D96A	2.8	4.5	2.2 ± 0.34	1.0 ± 0.02
C6.5V97A	3.0	4.8	3.1 ± 0.62	1.5 ± 0.02
C6.5G98A	19.8	31.7	4.1 ± 0.71	13 ± 0.55
C6.5Y99A	3.7	5.9	9.0 ± 0.17	5.3 ± 0.07
C6.5C100S/C100eS	17.5	28.0	5.0 ± 0.25	13.8 ± 0.71
C6.5S100aA	1.8	2.8	4.7 ± 0.55	1.3 ± 0.04
C6.5S100bA	2.9	4.7	3.4 ± 0.49	1.6 ± 0.07
C6.5S100cA	1.5	2.4	4.5 ± 0.62	1.1 ± 0.03
C6.5N100dA	1.8	2.9	4.1 ± 0.34	1.2 ± 0.05
C6.5K100gA	0.6	0.98	4.3 ± 0.31	0.42 ± 0.01
C6.5W100hA	NB	NB	NB	NB
C6.5P100iA	0.6	1.0	10.5 ± 0.12	1.1 ± 0.02
C6.5E100jA	NB	NB	NB	NB
C6.5Y100kA	101.0	161.6	0.73 ± 0.07	11.8 ± 0.25
C6.5F100lA	28.4	45.4	1.1 ± 0.13	5.0 ± 0.06
C6.5Q101A	0.5	0.82	12.0 ± 0.02	0.98 ± 0.02
C6.5H102A	1.2	1.9	5.9 ± 0.57	1.1 ± 0.02

Amino acid residues 95 to 99, 100a to 100d, and 100g to 102 of C6.5 V_H CDR3 were mutated to alanine using site-directed mutagenesis. Cysteine residues, C100 and C100e, were simultaneously mutated to serine. k_{on} and k_{off} were measured by SPR in a BIAcore, and the K_d calculated. Numbering is according to Kabat *et al.* (1991). NB, no binding.

(library D). The mutant phage antibody libraries were designated C6VHCDR3 libraries A, B, C, and D. PCR screening revealed that 30 of 30 randomly selected colonies from each library had full length insert and diversity was confirmed by sequencing ten unselected clones from each library (results not shown). Prior to selection, the percentage of clones expressing scFv which bound c-erbB-2 ECD by ELISA was 1% for C6VHCDR3 library A, 57%, library B, 2% library C, and 3% library D. The C6VHCDR3 libraries A, B, C, and D were selected on biotinylated c-erbB-2 ECD as described above and by Schier *et al.* (1996b), but using lower antigen concentration. The first round of selection was performed using 5.0×10^{-9} M c-erbB-2 ECD, tenfold lower than for the first round of selection of the C6VLCDR3 library. This concentration was chosen because the parental scFv for these libraries (C6ML3-9) had a greater than tenfold lower K_d than the parental clone for the C6VLCDR3 library (C6.5). Biotinylated c-erbB-2 ECD concentration was then decreased 100-fold for the second round of selection (5.0×10^{-11} M) and tenfold for the third and fourth rounds (5.0×10^{-12} M and 5.0×10^{-13} M). As for the C6VLCDR3 library, the rate of binding of polyclonal phage was measured in a BIAcore to determine the antigen concentration used for the subsequent round of selection (Schier & Marks, 1996).

Characterization of mutant scFv

After four rounds of selection, positive clones were identified by ELISA and at least 24 scFv from the fourth round of selection were ranked by k_{off} using SPR in a BIAcore. The ten scFv with the lowest k_{off} from C6VHCDR3 libraries A, C, and D were sequenced. Due to the diversity of isolated scFv in C6VHCDR3 library B, 48 scFv were ranked by k_{off} using SPR, and 22 clones with the lowest k_{off} were sequenced. scFv were purified by IMAC, followed by gel filtration to remove any dimeric or aggregated scFv. The k_{on} and k_{off} were determined by BIAcore and the K_d calculated. Very different results were obtained from the four libraries with respect to the number of higher affinity scFv isolated, and the value of the highest affinity scFv. The best results were obtained from library B (Table 4). Fifteen scFv were isolated with a K_d lower than wt C6ML3-9 and no wt sequences were observed. The best scFv (C6MH3-B47) had a $K_d = 1.1 \times 10^{-10}$ M, ninefold lower than C6ML3-9 and 145-fold lower than C6.5. The k_{off} of this scFv was 7.5×10^{-5} s^{-1}, tenfold lower than C6ML3-9 and 84-fold lower than C6.5. While a wide range of sequences was observed, a subset of scFv had the consensus sequence TDRT (first eight scFv, Table 4, library B). The consensus sequence is identical with the sequence of C6MH3-B1, which is the scFv with the lowest k_{off} (6.0×10^{-5} s^{-1}). Five scFv were isolated that had a k_{off} 2.5 to 3.75-fold lower than C6ML3-9, however expression levels were too low to obtain adequate purified scFv for measurement of the K_d (last five sequences, Table 4, library B).

The next best results were obtained from library D (Table 4, library D). Five higher affinity scFv were isolated, with the best having a K_d sevenfold higher than wt C6ML3-9. An additional scFv was isolated that had a k_{off} lower than wt scFv, however the expression level was too low to obtain adequate purified scFv for measurement of the K_d (last sequence, Table 4, library D). There was selection for a consensus mutation of Y100kW and replace-

ment of F100I with hydrophobic methionine or leucine.

No higher affinity scFv were isolated from either the A or C libraries. From library A, 8/10 scFv were wt, with one higher affinity scFv, a contaminant from library B. A single mutant scFv with the conservative replacement of Y99F had an apparent k_{off} 2.5 times lower than wt, but expression levels were too low to obtain adequate purified scFv to measure the K_d. From library C, 8/10 scFv were wt scFv, with one higher affinity scFv having mutations located in the V_H and V_L genes, but not in the region intentionally mutated. The isolated mutant scFv K100gV had a K_d 2.7-fold lower than wt (k_{off} 3.8-fold lower than C6ML3-9).

Ability of alanine scanning to identify residues which modulated affinity

Residue E100j, the only residue that when converted to alanine had no detectable binding, was 100% conserved. Otherwise, there was no correlation between the frequency with which the wt amino acid was recovered and the extent to which binding was reduced by substitution to alanine (Figure 2). Similarly, there was no correlation between residues shown to modulate affinity by alanine scanning and mutations exhibiting improved binding. This is clear when comparing the results obtained from library B (where no alanine mutant had more than a 2.9-fold increase in K_d) and library D (where K_d was markedly increased for two alanine mutants, Y100kA and F1001A). Despite the different alanine scan results, both libraries yielded similar nine- and sevenfold increases in affinity. This result appears to be different from the results of Lowman & Wells (1993), who found a mild ($R^2 = 0.71$) positive correlation between the frequency with which the wt amino acid was recovered from a phage library of human growth hormone mutants and the extent to which binding was reduced by alanine scanning.

Table 4. Sequences, affinities and binding kinetics of scFv isolated from heavy chain CDR3 libraries A, B, C, and D

Clone	F	V_H CDR3 sequence	Other mutations	K_d (10^{-10} M)	k_{on} (10^5 M^{-1} s^{-1})	k_{off} (10^{-4} s^{-1})
C6.5	—	HDVGYCSSSNCAKWPEYFQH		160.0	4.0 ± 0.20	63.0 ± 0.60
V_H CDR3 library A						
C6ML3-9 (wt)	8	-DVGY----------		10.0	7.6 ± 0.20	7.6 ± 0.30
C6MH3-A2	1	----F----------		nd	nd	2.9 ± 0.03[a]
C6MH3-A3	1	---------D----------		2.5	9.9 ± 0.52	2.5 ± 0.47
V_H CDR3 library B						
C6ML3-9 (wt)	0	------SSSN----------		10.0	7.6 ± 0.20	7.6 ± 0.30
C6MH3-B47	1	------TDRS----------		1.1	6.7 ± 0.63	0.75 ± 0.04
C6MH3-B1	1	------TDRT----------		1.2	5.0 ± 0.24	0.60 ± 0.06
C6MH3-B39	1	------TDPT----------		1.8	10.7 ± 0.84	1.9 ± 0.29
C6MH3-B48	1	------TDPS----------		2.3	5.6 ± 0.35	1.3 ± 0.01
C6MH3-B11	1	-------DRS----------		3.0	7.7 ± 0.41	2.3 ± 0.08
C6MH3-B5	1	------TDAT----------		3.4	6.8 ± 0.39	2.3 ± 0.07
C6MH3-B41	1	------TDRP----------		5.3	5.1 ± 0.27	2.7 ± 0.02
C6MH3-B2	1	------TDPR----------		5.8	5.5 ± 0.38	3.2 ± 0.08
C6MH3-B20	1	-------PAR----------		1.4	11.3 ± 1.29	1.6 ± 0.36
C6MH3-B16	1	------ADVR----------		2.0	8.0 ± 0.48	1.6 ± 0.40
C6MH3-B25	2	------LTTR----------		2.3	8.3 ± 0.54	1.9 ± 0.32
C6MH3-B21	1	------TTPL----------		2.6	9.1 ± 0.57	2.4 ± 0.23
C6MH3-B27	1	------KN-R----------		4.7	8.5 ± 0.36	4.0 ± 0.47
C6MH3-B9	2	------KTAA----------		4.6	7.2 ± 0.36	3.3 ± 0.43
C6MH3-B15	1	------E--R----------		5.9	5.1 ± 0.49	3.0 ± 0.06
C6MH3-B34	1	------QTDR----------	VL Q1R	nd	nd	2.0 ± 0.04[a]
C6MH3-B43	1	------EDYT----------		nd	nd	2.6 ± 0.04[a]
C6MH3-B46	1	------TTPR----------	VH K23Q VH V76G	nd	nd	2.8 ± 0.03[a]
C6MH3-B33	1	-------DQT----------		nd	nd	2.8 ± 0.04[a]
C6MH3-B31	1	------DDYT----------	VL P7L	nd	nd	2.9 ± 0.04[a]
V_H CDR3 library C						
C6ML3-9 (wt)	8	----------AKWPE----		10.0	7.6 ± 0.20	7.6 ± 0.30
C6MH3-C4	1	----------V--------		3.7	5.4 ± 0.94	2.0 ± 0.21
C6MH3-C3	1	----------	VH G15E VL N54D	6.5	4.9 ± 0.57	3.2 ± 0.01
V_H CDR3 library D						
C6ML3-9 (wt)	4	----------YFQH		10.0	7.6 ± 0.20	7.6 ± 0.30
C6MH3-D2	1	----------WLGV		1.4	8.3 ± 0.38	1.2 ± 0.02
C6MH3-D3	1	----------WLDN		2.7	7.4 ± 0.35	2.0 ± 0.25
C6MH3-D6	1	----------WMYP		3.5	5.2 ± 0.18	1.8 ± 0.01
C6MH3-D5	1	----------WM-M		5.8	3.6 ± 0.21	2.1 ± 0.02
C6MH3-D1	1	----------WLHV		7.5	3.6 ± 0.04	2.7 ± 0.04
C6MH3-D7	1	----------WQDP	VL N54S	nd	nd	3.1 ± 0.09[a]

k_{on} and k_{off} were determined in a BIAcore using purified scFv, and K_d calculated. Dashes indicate sequence identity. Mutations arising from PCR error and located outside V_H CDR3 are listed under the heading other mutations. F, frequency of isolated scFv.
[a] k_{off} determined from unpurified scFv samples.

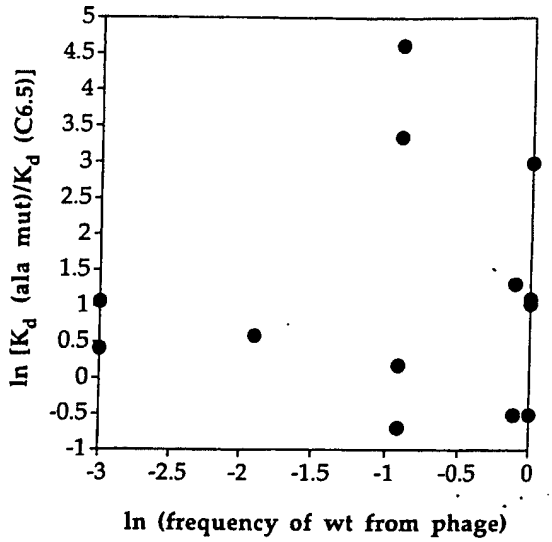


Figure 2. Relationship between the frequency with which a wild-type amino acid is recovered from V_H CDR3 phage antibody libraries and the extent to which binding is reduced when the residue is converted to alanine. The relative change in binding affinity from wt for alanine substitutions is plotted as ln [K_d (Ala mutant)/K_d (C6.5)]. Data for residue N100d could not be used because the wt residue was not recovered at this position. Data for residue E100j could not be used because no binding could be detected. No correlation exists between the frequency with which a wt amino acid is recovered from V_H CDR3 phage antibody libraries and the extent to which binding is reduced when the residue is converted to alanine.

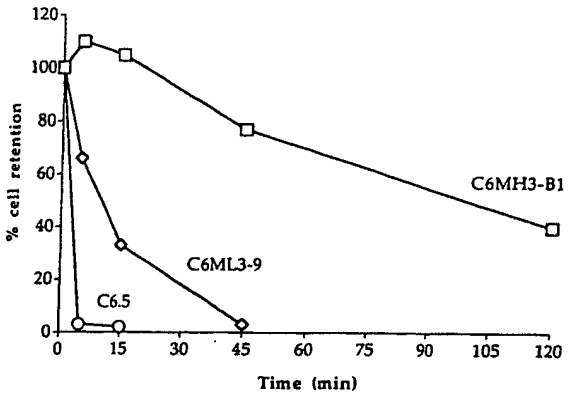


Figure 3. Cell surface retention of wild-type and mutant scFv. Retention of biotinylated C6.5 (k_{off} = 6.3 × 10 − 3 s^{-1}) on the surface of c-erbB-2 expressing SK-OV-3 cells was determined by FACS and compared to the retention of C6ML3-9 ($k_{off} = 7.6 \times 10^{-4}\ s^{-1}$) and C6MH3-B1 ($k_{off} = 6.0 \times 10^{-5}\ s^{-1}$). The half life ($t_{1/2}$) of the scFv on the cell surface was much less than five minutes for C6.5, 11 minutes for C6ML3-9, and 102 minutes for C6MH3-B1. These values agree closely with the $t_{1/2}$ calculated from the k_{off} as determined by SPR in a BIAcore (1.6 minutes for C6.5, 13 minutes for C6ML3-9, and 135 minutes for C6MH3-B1; Figure 4).

In addition, their largest improvements in affinity were for those residues shown by alanine scanning to significantly affect binding. The reason for the different results is unclear, however in two of our V_H CDR3 libraries where alanine scanning indicated a significant effect on binding (library A and C), expression levels of mutants were generally low. This could have affected the selection results.

Correlation between affinity and cell surface retention of scFv

The retention of biotinylated C6.5, C6ML3-9, and C6MH3-B1 scFv on the surface of SK-OV-3 cells expressing c-erbB-2 was determined, both to verify the observed differences in k_{off}, and to confirm that the antigen as presented in the BIAcore had biologic significance. The half life ($t_{1/2}$) of the scFv on the cell surface was much less than five minutes for C6.5, 11 minutes for C6ML3-9, and 102 minutes for C6MH3-B1 (Figure 3). These values agree closely with the $t_{1/2}$ calculated from the k_{off} as determined by SPR in a BIAcore (1.6 minutes for C6.5, 13 minutes for C6ML3-9, and 135 minutes for C6MH3-B1; Figure 4). The anti-digoxin scFv 26-10 (Huston *et al.*, 1988) was used as negative control, and no binding to c-erbB-2 ECD in a BIAcore or to c-erbB-2 on SK-OV-3 cells was observed.

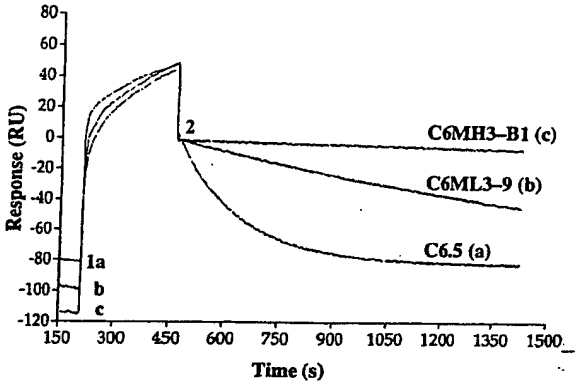


Figure 4. Dissociation of wild-type and mutant scFv from a c-erbB-2 ECD coupled sensor chip in a BIAcore. Overlay plot of sensorgrams of the association and dissociation of C6.5, C6ML3-9, and C6MH3-B1 scFv to a c-erbB-2 coupled sensor chip in a BIAcore. The decrease in k_{off} which accounts for the majority of the increase in K_d correlates closely with the $t_{1/2}$ of retention on SK-OV-3 cells (Figure 3). Sensorgrams were positioned so that the amount (RU) of scFv bound at the beginning of dissociation was superimposed. For this Figure, 5.0×10^{-7}M scFv were injected over a BIAcore flow cell coated with 1000 RU of c-erbB-2 ECD under a constant flow of 25 µl/minute. scFv were allowed to dissociate for a time period of 15 minutes. (1) Injection of scFv and beginning of the association phase. (2) End of the association phase, beginning of the dissociation phase. a, C6.5; b, C6ML3-9; c, C6MH3-B1.

Table 5. Binding kinetics of scFv derived from C6.5 V_L CDR3, V_H CDR3 and light chain shuffled mutants

Clone	K_d (10^{-10} M)	k_{on} (10^5 s^{-1} M^{-1})	k_{off} (10^{-4} s^{-1})	K_d (parent)/ K_d (mut)	K_d (C6.5)/ K_d (mut)	$\Delta\Delta G_I$ (kcal/mol)
A. *Combined mutants: C6ML3-9 or C6ML3-12 with light chain shuffled C6L1*						
C6-9L1	3.3	9.2 ± 0.20	3.0 ± 0.40	3.0	49	+0.42
C6-12L1	1.9	6.7 ± 0.12	1.3 ± 0.32	8.4	84	−0.18
B. *Combined mutants: C6MH3-B1 or C6MH3-B47 with light chain shuffled C6L1*						
C6-B1L1	6.3	3.8 ± 0.19	2.4 ± 0.01	0.19	25	+0.43
C6-B47L1	6.0	3.0 ± 0.16	1.8 ± 0.01	0.18	27	+0.45
C. *Combined mutants: C6MH3-B1 or C6MH3-B47 with D library mutants*						
C6-B1D1	0.32	4.7 ± 0.31	0.15 ± 0.005	3.8	500	−0.61
C6-B1D2	0.15	6.9 ± 0.42	0.10 ± 0.014	8.0	1067	−0.07
C6-B1D3	0.13	6.4 ± 0.20	0.08 ± 0.002	9.2	1231	−0.53
C6-B1D5	0.35	5.1 ± 0.36	0.18 ± 0.001	3.4	457	−0.40
C6-B1D6	0:32	4.1 ± 0.17	0.13 ± 0.002	3.8	500	−0.16
C6-B47D1	0.68	7.1 ± 0.95	0.48 ± 0.001	1.6	235	−0.11
C6-B47D2	0.44	9.8 ± 0.72	0.43 ± 0.001	2.5	364	+0.62
C6-B47D3	0.48	6.6 ± 0.26	0.32 ± 0.001	2.3	333	+0.29
C6-B47D5	0.63	6.2 ± 0.31	0.39 ± 0.002	1.7	254	−0.01
C6-B47D6	0.51	5.9 ± 0.30	0.30 ± 0.001	2.2	314	+0.17

A, Mutants produced by combining the V_L CDR3 of C6ML3-9 or C6ML3-12 with the V_L gene of the C6.5 light chain shuffled scFv, C6L1 (Schier *et al.*, 1996b). B, Mutants produced by introducing mutations in FR1 to FR3 of C6L1 light chain into C6MH3-B1 or C6MH3-B47. C, Mutants obtained by combining mutations of C6MH3-B1 or C6MH3-B47 with mutations from D library clones (D1, D2, D3, D5, D6). k_{on}, and k_{off} were measured by SPR in a BIAcore, and the K_d calculated. $\Delta\Delta G_I$ were calculated as by Lowman & Wells (1993), and Cunningham & Wells (1993).

Effects on binding kinetics by combining mutations from high affinity scFv

To further increase affinity, the sequences of the two highest affinity scFv obtained from the V_H CDR3B library (C6MH3-B1 or C6MH3-B47) were combined with the sequences of scFv isolated from the C6VHCDR3D library (C6MH3-D1, -D2, -D3, -D5, or -D6). An increase in affinity from wt was obtained for all these combinations, yielding an scFv (C6-BlD3) that had a 1230-fold lower K_d than wt C6.5 (Table 5). The extent of additivity varied considerably, however, and could not be predicted from the parental k_{on}, k_{off}, or K_d. In some combinations, cooperativity was observed, with a negative $\Delta\Delta G_I$. Additional combinations were made between a previously described light chain shuffled C6.5 mutant (C6L1, sixfold decreased K_d; Schier *et al.*, 1996b) and one of two V_L CDR3 mutants (C6ML3-9 and C6ML3-12). These combinations yielded scFv with 49 and 84-fold improved affinity (Table 5). Introducing the same rearranged V_L gene into the highest affinity V_H CDR3 mutants (C6MH3-B1 or C6MH3-B47) resulted in decreased affinity compared to C6MH3-B1 (Table 5). A similar effect was described in previous work (Schier *et al.*, 1996b) when rearranged V_L and V_H genes from high affinity chain shuffled scFv obtained from parallel selection were combined.

Discussion

Ultra-high affinity scFv were engineered by diversifying the CDRs that comprise the center of the antibody combining site. Sequential diversification of V_L and V_H CDR3 yielded scFv with up to a 145-fold increase in affinity ($K_d = 1.1 \times 10^{-10}$ M). Combination of these mutations with independently selected mutations located elsewhere in V_H CDR3 yielded an additional ninefold increase in affinity ($K_d = 1.3 \times 10^{-11}$ M). The scFv were produced without any immunization and have higher affinity than any antibody fragments engineered *in vitro*. The results illustrate the power of diversity libraries and phage display to produce antibody fragments with affinities rarely achieved by immunization (Foote & Eisen, 1995) and have important implications for the design of mutant phage antibody libraries. Moreover, the availability of such high affinity antibody fragments may have important consequences for antibody based tumor targeting.

Accuracy of affinity measurements

The validity of our results depends on the accuracy of the measured affinities, which were calculated from k_{on} and k_{off} determined by BIAcore. To verify that differences in k_{on} were not due to differences in the immunoreactivity of the purified scFv, the concentration of functional scFv was determined by measuring the binding rate to c-erbB-2 ECD under mass transport limited conditions (Karlsson *et al.*, 1993; Schier *et al.*, 1996b). Since increases in affinity were largely due to a decrease in k_{off}, precautions were taken to avoid the introduction of artifact into these measurements (Nieba *et al.*, 1996). Purified scFv were gel filtered immediately prior to k_{off} measurement, to avoid avidity from dimeric or aggregated antibody fragment (Schier *et al.*, 1996b), and analytical gel filtration was performed after measurement of k_{off}

to confirm the absence of aggregated material. To minimize the probability of rebinding, k_{off} was measured using a high flow rate and a scFv concentration that resulted in near saturation of the chip surface. The amount of c-erbB-2 ECD coupled to the chip surface was the lowest amount that gave an adequate binding response (100 to 150 RU) for accurate kinetic measurement. In our experience, using the minimal amount of coupled antigen is the single most important parameter for preventing rebinding. Using these experimental conditions, we were unable to detect any evidence of rebinding when k_{off} was measured in the presence of 5×10^{-7} M c-erbB-2 ECD in the running buffer. Furthermore, the affinity of C6.5 previously determined by Scatchard after radiolabeling (2.0×10^{-8} M, Schier *et al.*, 1995) agrees closely with the value determined by BIAcore (1.6×10^{-8} M). The k_{off} of CDR3 mutants determined by cell surface retention of biotinylated scFv also agrees closely with k_{off} measured by BIAcore. Engineering further increases in affinity is likely to require a different technique for affinity measurement, since the binding kinetics of the highest affinity scFv are near the limit of measurement using SPR in a BIAcore ($k_{on} > 10^6$ M^{-1} s^{-1} and $k_{off} < 10^{-5}$ s^{-1}; Malmqvist, 1993). Determining k_{off} below 10^{-5} s^{-1} is difficult due to the small amount of analyte dissociating (1% in 17 minutes), the background noise, and disturbances from the pumps and valves of the flow system. Determination of K_d in the BIAcore using competition experiments (Nieba *et al.*, 1996) will also be limited by instrument sensitivity.

Design of mutant antibody libraries

When designing a mutant phage antibody library, decisions must be made as to how and where to introduce mutations. Mutations can be randomly introduced, using either chain shuffling (Clackson *et al.*, 1991; Marks *et al.*, 1992), error prone PCR (Hawkins *et al.*, 1992), or mutator strains (Low *et al.*, 1996), thus apparently mimicking the process of somatic hypermutation. These approaches have yielded large increases in affinity for hapten antigens (>100-fold; Low *et al.*, 1996; Marks *et al.*, 1992), but results with protein binding antibody fragments have been more modest (<tenfold; Hawkins *et al.*, 1992; Schier *et al.*, 1996b). Moreover, the relatively random distribution of mutations in higher affinity clones provides little useful information as to where to direct additional mutations. Alternatively, knowledge of the general structure of the Fv fragment and its complexes with antigen can be used to direct mutagenesis to the CDRs that form the contact interface between antibody and antigen.

Targeting mutations to the CDRs has previously been shown to be an effective technique for increasing antibody affinity. Yang *et al.* (1995) increased the affinity of an anti-HIV gp120 Fab 420-fold ($K_d = 1.5 \times 10^{-11}$ M) by mutating four CDRs in five libraries and combining independently selected mutations. We achieved three times that increase in affinity by mutating a much smaller portion of the antibody combining site contained within only two CDRs. Our results may be partly due to the stringent selection conditions used and the techniques used to monitor selections and screen for higher affinity scFv without the need for purification. However, the results also suggest that focusing mutations in V_H and V_L CDR3 may be a more efficient means to increase affinity.

Directing mutations into V_H and V_L CDR3 to increase affinity may initially seem at odds with studies on antibody structure and function. Although 15 to 22 amino acids located in loops within the CDRs typically contact antigen (Davies *et al.*, 1990), free energy calculations and mutational analysis indicate that only a small subset of the contact residues contribute the majority of the binding energy (Hawkins *et al.*, 1993; Kelley & O'Connell, 1993; Novotny *et al.*, 1989). The high energy contact residues are more frequently located in the center of the antibody combining sites in the V_H and V_L CDR3s. Thus, mutation of V_H and V_L CDR3 is more likely to destroy high affinity contacts than mutation of other CDRs. However, these residues will be recreated, albeit at low frequency, given an adequate library size for the number of residues randomized. Mutant residues could increase affinity by introducing new contact residues (Alzari *et al.*, 1990) or by replacing low affinity (Novotny *et al.*, 1989; Kelley & O'Connell, 1993) or "repulsive" contact residues (Novotny *et al.*, 1989) with contact residues with more favorable energetics. It appears, however, that many mutations introduced either by somatic hypermutation *in vivo* (Sharon, 1990) or mutagenesis *in vitro* (Hawkins *et al.*, 1993) exert their affect on affinity indirectly, in many instances by precisely positioning the side-chains of contact residues for optimal electrostatic, hydrogen bonding, and van der Waals interactions (Mian *et al.*, 1991). Mutation of non-contact CDR residues located close to high energy contact CDR residues may be more likely to exert this indirect effect. The importance of the CDR3s as sites for mutagenesis is also supported by the work of Yang *et al.* (1995) who created separate libraries of V_L CDR1, V_L CDR3, V_H CDR1, and V_H CDR3 mutants. The largest increases in affinity were 7.9 and 7.7-fold from sequential mutation of two separate regions of V_H CDR3, resulting in a 63-fold increase in affinity over wt Fab. Mutation of V_L CDR3 resulted in the next largest increase in affinity over wt (5.6-fold).

Directing mutations into V_H and V_L CDR3 to increase affinity may also appear to be different from the locations where mutations are directed and accumulate during somatic hypermutation *in vivo*. Germline diversity is greatest in the center of the antibody combining site (Tomlinson *et al.*, 1996), particularly V_H and V_L CDR3, where tremendous sequence diversity is generated by recombination, N segment addition, and joining diversity. Somatic hypermutation extends sequence

diversity to CDR residues located more peripherally in the antibody combining site (Tomlinson *et al.*, 1996). On closer inspection, however, striking similarities exist between our results and somatic hypermutation *in vivo*. Nucleotide substitutions *in vivo* are not targeted randomly, but rather occur at specific sequence hotspots intrinsic to the mutational process, for example at serine residues encoded by the nucleotides AGY but not at serine residues encoded by TCN (Betz *et al.*, 1993; Reynaud *et al.*, 1995). Serine residues encoded by AGY have previously been shown to predominate over those encoded by TCN in CDR1 and two of the V_H domain and CDR1 of the V_κ domain, but not in the frameworks (Betz *et al.*, 1993; Wagner *et al.*, 1995). The germline gene segments encoding the V_κ and V_λ CDR3s are similarly biased to contain a high proportion of AGY serine (Figure 5). The CDR3 AGY/TCN ratio is 9.7 for the V_κ gene segments and 5.4 for the V_λ gene segments. These values are comparable to the AGY/TCN ratios observed for V_H CDR1 (20.3), V_H CDR2 (2.28), and V_κ CDR1 (8.75) and are greater than values observed for framework residues (0.3 to 0.67; Wagner *et al.*, 1995). Thus the sequences encoding the V_L CDR3 have evolved to be targets of the somatic hypermutation machinery. Accordingly, an extensive analysis of the location of mutations introduced by somatic hypermutation into the germline V_κ genes identified CDR3 as a frequent site of mutagenesis (Tomlinson *et al.*, 1996). Within V_κ CDR3, residues 89 to 91 are most conserved, with the highest frequency of mutation observed at residue 93 (Tomlinson *et al.*, 1996). Similarly, during *in vitro* affinity maturation of the C6.5 V_λ CDR3, we observed conservation of residues 89 to 92, and 95b (this work) and 96 (Schier *et al.*, 1996a). In contrast, substitution occurred at residues 93 to 95a, with the highest frequency of mutation at residue 93. The location of mutations parallels exactly the distribution of AGY serine residues within the germline V_λ CDR3 genes (Figure 5).

Such a detailed analysis is more difficult for V_H CDR3, since germline D gene segment assignment is frequently not possible, and the V_H CDR3 loop cannot be accurately modeled. In 33 published D gene segments (Kabat *et al.*, 1991), the ratio of AGY/TCN serine residues is 2.5/1, a value closer to that observed for CDRs (2.3 to 20.3) than for framework residues (0.3 to 0.67). Thus portions of the V_H CDR3 also appear to be targets for somatic hypermutation *in vivo*. In the case of the C6.5, the sequence motif CSSSNC located within the V_H CDR3 is encoded by the germline D gene segment D1 (Kabat *et al.*, 1991) which encodes the sequence CSSTSC. In the germline gene, all three serine residues are encoded by AGY and in C6.5 these three residues were hotspots for substitutions which increased affinity. Moreover, an extremely wide range of amino acid residues were tolerated between the two cysteine residues, as evidenced by the alanine scan results and the fact that 57% of unselected scFv bound antigen. In contrast, only 1

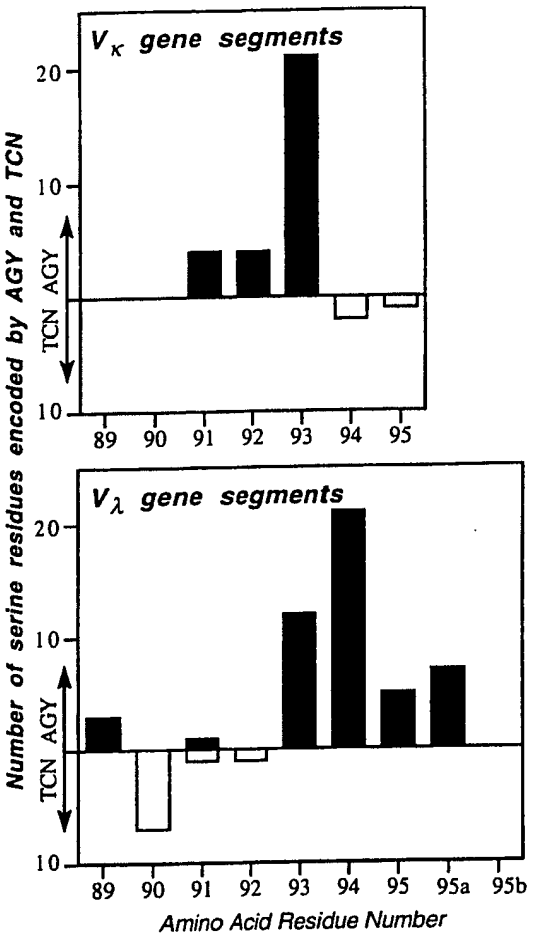


Figure 5. Frequency of serine encoded by AGY or TCN in the CDR3s of the V_κ and V_λ germline gene segments. The DNA sequences of 50 germline V_κ genes (Cox *et al.*, 1994) and 24 germline V_λ genes (Williams & Winter, 1993) were analyzed and the serine codon usage (AGY or TCN) plotted as a function of residue number within the CDR3 (residues 89 to 95 for V_κ and 89 to 95b for V_λ). The number of serine residues encoded by AGY are shown by filled bars and the number of serine residues encoded by TCN by open bars.

to 2% of unselected scFv from the three other V_H CDR3 libraries bound antigen. One possible explanation is that the disulfide bond formed by the cysteine residues stabilizes this region of the V_H CDR3. Regardless, five of the 33 D gene segments encode a CXXXXC or CXXXC (Kabat *et al.*, 1991) in the preferred reading frame (Yamada *et al.*, 1991; Abergel & Claverie, 1991) suggesting that the motif has useful properties as a component of the primary immune repertoire. If so, then use of similar motifs could prove useful in construction of semi-synthetic antibody libraries (Hoogenboom & Winter, 1992).

Modeling the location of the observed mutations in V_L CDR3 on the Fab structure KOL, where the V_λ is derived from the same germline gene as C6.5,

suggested a basis for the pattern of conservation and substitution we observed. The conserved residues appear to have a significant structural role in the variable domain, either in maintaining the main-chain conformation of the loop, or in packing on the V_H domain. In our model, the side-chain of W91 is buried at the V_H-V_L interface and by the long V_H CDR3 of KOL. In other structures, the side-chain of V_L91 is at least partially solvent accessible, and frequently contacts antigen (Mian *et al.*, 1991; Tomlinson *et al.*, 1996). In contrast to the extensive packing of the side-chains of conserved residues, all but one of the four most extensively substituted residues appear to have solvent accessible side-chains. A similar pattern was also observed during parsimonious mutagenesis of C6.5 V_L CDR1 and V_H CDR2 (Schier *et al.*, 1996a). Residues within the CDRs with structural roles were conserved, while substitutions were largely confined to residues with solvent accessible side-chains. Analysis of the location of substitutions in V_H and V_L genes mutated *in vivo* also indicates that residues involved in maintaining the main-chain conformation rarely undergo non-conservative substitution (Chothia *et al.*, 1992; Tomlinson *et al.*, 1995).

The previous analysis suggests a mutagenesis strategy for efficiently increasing antibody fragment affinity. Mutagenesis is directed into V_L and V_H CDR3 sequentially, as in this work, rather than by parallel evolution of the two CDRs. These two CDRs pack on each other, and mutations isolated in parallel are likely to not be additive (Yang *et al.*, 1995). Mutagenesis is initially directed into V_L CDR3 due to the limited main-chain conformations (Chothia & Lesk, 1987; Tomlinson *et al.*, 1995) and the ability to model the CDR on a homologous Fv or Fab structure. Modeling should be used to identify CDR residues that are likely to have a structural role, either in maintaining the main-chain conformation or in packing against the V_H domain. These residues are conserved, leaving at most four to five residues which can be completely randomized in a reasonably sized library (10^7 to 10^8 member). The highest affinity V_L CDR3 mutant is then used as a template for V_H CDR3 mutagenesis. Given the length of V_H CDR3s, it is likely that it will not be possible to sample the entire sequence space simultaneously. Instead residues are randomized four to five at a time, as in this work, and independently selected mutations combined. In this and previous work (Schier *et al.*, 1996a), we observed complete conservation of the four glycine and two tryptophan residues randomized. In the CDRs, glycine residues are typically key residues in turns, and the chemical properties of tryptophan make it a frequent structural or high energy contact residue (Mian *et al.*, 1991). Thus conservation of these two residues when randomizing V_H CDR3 should be considered, if sequence space is limiting. Since combination of independently selected mutations may not be additive, it may be more prudent to simultaneously scan all residues at a low mutation frequency (parsimonious mutagenesis) to identify residues that modulate affinity, and structural and functional residues that are conserved (Schier *et al.*, 1996a). Residues identified as modulating affinity would then be completely randomized in a second library. Alanine scanning appeared to be useful only to identify essential contact and structural residues, but not for predicting which residues would yield higher affinity when mutated. If necessary, affinity could be increased further by mutating the other CDRs. Particularly suitable might be V_H CDR1 and V_L CDR1. These CDRs appear to be more important in modulating affinity during *in vivo* affinity maturation, based both on the higher frequency of AGY serine in the germline genes (Wagner *et al.*, 1995), and a higher frequency of mutation in rearranged genes (Tomlinson *et al.*, 1996). Modeling should be performed to identify structural residues to be conserved, and residues with solvent accessible side-chains, which would be mutated.

Implications for antibody based tumor targeting

The availability of scFv with a range of affinities for a tumor antigen makes it possible to determine the effect of affinity on specific tumor retention. While it might appear obvious that retention should increase with increasing affinity, it has been proposed that a barrier effect exists, such that the higher affinity antibodies are trapped at the tumor edge (Fujimori *et al.*, 1990). We have examined the effect of C6.5 mutants with affinities ranging between 3.2×10^{-7} M (C6G98A) and 1.0×10^{-9} M (C6ML3-9) in *scid* mice bearing human SK-OV-3 tumors. The percentage injected dose/gram of tumor at 24 hours increased from 0.19% for C6G98A to 1.42% for C6ML3-9 and the tumor:blood ratios increased from 2.6 to 17.2 (unpublished results). Thus within the range of affinities studied, there was a significant increase in tumor retention. The magnitude of the 24 hour retention, however, is significantly less than the values observed for IgG (30% ID/g) in similar models. Although an affinity of 1.0×10^{-9} M is considered high, the k_{off} gives a predicted $t_{1/2}$ on the cell surface of only 13 minutes, much faster than the β elimination $t_{1/2}$ of the scFv from the mouse (2.5 hours). Thus tumor retention is largely dependent on the rate of clearance from the blood. The higher affinity scFv described in this paper have a k_{off} which provide a predicted $t_{1/2}$ on the cell surface of 24 hours, significantly longer than the clearance rate from blood. These very high affinity antibody fragments, with binding kinetics not previously available, offer the possibility of significantly greater quantitative tumor retention. The scFv could also be used as building blocks to create dimeric scFv, with yet higher apparent affinity due to avidity, and even greater tumor retention (Adams *et al.*, 1993).

Materials and Methods

Construction of phage antibody libraries

Mutant scFv phage antibody libraries were constructed based on the sequence of C6.5, a human scFv isolated from a non-immune phage antibody library which binds to the tumor antigen c-erbB-2 with a $K_d = 1.6 \times 10^{-8}$ M (Schier *et al.*, 1995). For construction of a library containing V_L CDR3 mutants, an oligonucleotide (V_L1; Table 1) was designed which partially randomized nine amino acid residues located in V_L CDR3 (Table 2). For the nine amino acids randomized, the ratio of nucleotides was chosen so that the frequency of wt amino acid was 49%. To create the library, C6.5 scFv DNA (10 ng) was amplified by PCR in 50 µl reactions containing 25 pmol LMB3 (Marks *et al.*, 1991), 25 pmol VL1, 250 µM dNTPs, 1.5 mM $MgCl_2$, and 1 µl (five units) *Taq* DNA polymerase (Promega) in the manufacturer's buffer. The reaction mixture was subjected to 30 cycles of amplification (94°C for 30 seconds, 50°C for 30 seconds and 72°C for one minute) using a Hybaid OmniGene cycler. To introduce a *Not*I restriction site at the 3' end of the scFv gene repertoire, the PCR fragment (850 bp) was gel purified and reamplified using the primers LMB3 and V_L2 (Table 1). The PCR product was purified, digested with *Sfi*I and *Not*I, and ligated into pCANTAB5E (Pharmacia) digested with *Sfi*I and *Not*I. Ligation mixtures were purified as previously described (Schier *et al.*, 1996b) and aliquots electroporated (Dower *et al.*, 1988) into 50 µl *E. coli* TG1 (Gibson, 1984). Cells were grown in 1 ml SOC (Sambrook *et al.*, 1990) for 30 minutes and then plated on TYE (Miller, 1972) medium containing 100 µg ampicillin/ml and 1% (w/v) glucose (TYE-AMP-Glc). Colonies were scraped off the plates into 5 ml of 2 × TY broth (Miller, 1972) containing 100 µg ampicillin/ml, 1% (w/v) glucose (2 × TY-AMP-Glc) and 15% (v/v) glycerol for storage at –70°C. The cloning efficiency and diversity of libraries was determined by PCR screening (Gussow & Clackson, 1989) exactly as described by Marks *et al.* (1991) and by DNA sequencing (Sanger *et al.*, 1977) The mutant phage antibody library was designated C6VLCDR3.

Four libraries of V_H CDR3 mutants were constructed. For construction of each V_H CDR3 library, oligonucleotides (VHA, VHB, VHC, and VHD; Table 1) were designed which completely randomized four amino acid residues located in V_H CDR3 (amino acid residues 96 to 99, library A; residues 100a to 100d, library B; residues 100f, 100g, 100i, and 100j, library C; and residues 100k to 102, library D; Table 3). To create the libraries, DNA encoding the V_H gene of C6.5 scFv DNA (10 ng) was amplified by PCR in 50 µl reactions containing 25 pmol LMB3 (Marks *et al.*, 1991) and 25 pmol of either VHA, VHB, VHC, or VHD exactly as described above. The resulting PCR fragments were designated VHA1, VHB1, VHC1, and VHD1, based on the mutagenic oligonucleotide used for amplification. In four separate PCR reactions, DNA encoding the light chain, scFv linker, V_H framework 4 (FR4), and a portion of V_H CDR3 of C6ML3-9 was amplified by PCR as described above using the primers C6hisnot and either RVHA, RVHB, RVHC, or RVHD (Table 1). These amplifications yielded PCR fragments VHA2, VHB2, VHC2, and VHD2. The 5' ends of primers RVHA, RVHB, RVHC, and RVHD were designed to be complementary to the 5' ends of primers VHA, VHB, VHC, and VHD, respectively. This complementarity permits joining of the VH1 and VH2 PCR fragments together to create a full length scFv gene repertoire using splicing by overlap extension. To create the mutant scFv gene repertoires, 200 ng of each PCR fragment (VHA1 and VHA2, VHB1 and VHB2, VHC1 and VHC2, or VHD1 and VHD2) were combined in 50 µl PCR reaction mixtures (as described above) and cycled seven times to join the fragments (94°C for 30 seconds, 60°C for five seconds, 40°C for five seconds (RAMP: five seconds), 72°C for one minute). After seven cycles, outer primers LMB3 and C6hisnot were added and the mixtures amplified for 30 cycles (94°C for 30 seconds, 50°C for 30 seconds, 72°C for one minute). The PCR products were purified as described above, digested with *Sfi*I and *Not*I, and separately ligated into pCANTAB5E (Pharmacia) digested with *Sfi*I and *Not*I. The four ligation mixtures were purified as described above and electroporated into 50 µl *E. coli* TG1. Transformed cells were grown and plated, and libraries characterized and stored, as described above. The mutant phage antibody libraries were designated C6VHCDR3A, C6VHCDR3B, C6VHCDR3C, and C6VHCDR3D.

Preparation of phage and selection of phage antibody libraries

Preparation of phage for selection was performed exactly as described by Schier *et al.* (1996b). Phage particles were purified and concentrated by two PEG precipitations (Sambrook *et al.*, 1990), resuspended in 5 ml phosphate buffered saline (25 mM NaH_2PO_4, 125 mM NaCl, pH 7.0; PBS) and filtered through a 0.45 µm filter. All libraries were selected using biotinylated c-erbB-2 ECD and streptavidin-coated paramagnetic beads M280 (Dynal) as described by Schier *et al.* (1996b). For selection of the C6VLCDR3 library, c-erbB-2 ECD concentrations of 4.0×10^{-8} M, 1.0×10^{-9} M, 1.0×10^{-10} M, and 1.0×10^{-11} M were used for selection rounds 1, 2, 3, and 4, respectively. The mixture of phage and antigen was gently rotated for one hour at room temperature and phage bound to biotinylated antigen captured using 100 µl (round 1) or 50 µl (rounds 2, 3, and 4) of streptavidin-coated M280 magnetic beads. After capture of phage, Dynabeads were washed a total of ten times (three times in PBS containing 0.05% Tween 20 (TPBS), twice in TPBS containing 2% skimmed milk powder (2% MTPBS), twice in PBS, once in 2% MPBS, and twice in PBS) using a Dynal magnetic particle concentrator. The Dynabeads were resuspended in 1 ml PBS, and 300 µl were used to infect 10 ml log phase *E. coli* TG1 which were plated on TYE-AMP-Glc plates. For selection of the C6VHCDR3 libraries, c-erbB-2 ECD concentrations of 5.0×10^{-9} M, 5.0×10^{-11} M, 5.0×10^{-12} M, and 5.0×10^{-13} M were used for selection rounds 1, 2, 3, and 4, respectively and the phage captured by incubating with 50 µl of Dynabeads for five minutes. The washing protocol was altered to select for scFv with the lowest k_{off} (Hawkins *et al.*, 1992). Dynabeads with bound phage were initially subjected to five rapid washes (4 × TPBS, 1 × MPBS) followed by six 30 minute incubations in one of three washing buffer (2 × TPBS, 2 × MPBS, 2 × PBS) containing 1.0×10^{-7} M c-erbB-2 ECD. Bound phage were eluted from the Dynabeads by sequential incubation with 100 µl of 4 M $MgCl_2$ for 15 minutes followed by 100 µl of 100 mM HCl for five minutes. Eluates were combined and neutralized with 1.5 ml of 1 M Tris HCl (pH 7.5) and one third of the eluate used to infect log phase *E. coli* TG1.

Initial scFv characterization

Initial analysis of selected scFv clones for binding to c-erbB-2 ECD was determined by phage ELISA. To prepare phage for ELISA, single ampicillin-resistant colonies were transferred into microtiter plate wells containing 100 µl 2 × TY-AMP, 0.1% glucose and grown for three hours at 37°C to an A_{600} of approximately 0.5. VCSM13 helper phage (2.5×10^8 phage) were added to each well, and the cells incubated for one hour at 37°C. Kanamycin was then added to each well to a final concentration of 25 µg/ml and the bacteria grown overnight at 37°C. Supernatants containing phage were used for ELISA. For ELISA, Immunolon 4 plates (Dynatech) were incubated overnight at 4°C with ImmunoPure avidin (10 µg/ml in PBS; Pierce). After washing three times with PBS to remove unbound avidin, wells were incubated with biotinylated c-erbB-2 ECD as described by Schier *et al.* (1995). Binding of scFv phage to c-erbB-2 ECD was detected with peroxidase-conjugated anti-M13 antibody (Pharmacia) and ABTS (Sigma) as substrate. Selected binders were further characterized by DNA sequencing of the V_H and V_L genes (Sanger *et al.*, 1977).

Ranking of scFv by k_{off} was performed using SPR in a BIAcore as described by Schier *et al.* (1996b). Briefly, 10 ml cultures of 24 ELISA positive clones from the third and fourth round of selection were grown to an A_{600} of approximately 0.8, scFv expression induced (De Bellis & Schwartz, 1990) and the culture grown overnight at 25°C. scFv were harvested from the periplasm (Breitling *et al.*, 1991), and the periplasmic fraction dialyzed for 48 hours against Hepes buffered saline (10 mM Hepes, 150 mM NaCl, pH 7.4; HBS). In a BIAcore flow cell, approximately 1400 resonance units (RU) of c-erbB-2 ECD were coupled to a CM5 sensor chip (Schier *et al.*, 1996b) using NHS-EDC chemistry (Johnsson *et al.*, 1991). Association and dissociation of undiluted scFv in the periplasmic fraction were measured under a constant flow of 5 µl/minute and HBS as running buffer. An apparent k_{off} was determined from the dissociation part of the sensorgram for each scFv analyzed (Karlsson *et al.*, 1991). The flow cell was regenerated between samples using sequential injections of 4 M $MgCl_2$ and 100 mM triethylamine without significant change in the sensorgram baseline after analysis of more than 100 samples.

Subcloning, expression and purification of scFv

To facilitate purification of scFv selected from the C6VLCDR3 library, the scFv genes were subcloned (Schier *et al.*, 1995) into the expression vector pUC 119 Sfi-NotmycHis, which results in the addition of a hexa-histidine tag at the C-terminal end of the scFv. The scFv selected from the C6VHCDR3 library already have a C-terminal hexa-histidine tag and therefore could be purified without subcloning. Cultures (500 ml) of *E. coli* TG1 harboring one of the C6.5 mutant phagemids were grown, expression of scFv induced (De Bellis & Schwartz, 1990), and the culture grown at 25°C overnight. scFv were harvested from the periplasm (Breitling *et al.*, 1991), dialyzed overnight at 4°C against eight liters of IMAC loading buffer (50 mM sodium phosphate (pH 7.5), 500 mM NaCl, 20 mM imidazole) and then filtered through a 0.2 µm filter.

scFv was purified by IMAC (Hochuli *et al.*, 1988) exactly as described by Schier *et al.* (1995). To separate monomeric, dimeric and aggregated scFv, samples were concentrated to a volume <1 ml in a Centricon 10 (Amicon) and fractionated on a Superdex 75 column using a running buffer of HBS. The purity of the final preparation was evaluated by assaying an aliquot by SDS-PAGE. Protein bands were detected by Coomassie staining. The concentration was determined spectrophotometrically, assuming an A_{280} nm of 1.0 corresponds to an scFv concentration of 0.7 mg/ml.

Measurement of affinity and binding kinetics

The K_d of scFv were determined using SPR in a BIAcore (Schier *et al.*, 1996b). In a BIAcore flow cell, approximately 1400 RU of c-erbB-2 ECD (90 kDa, McCartney *et al.*, 1995) were coupled to a CM5 sensor chip (Johnsson *et al.*, 1991). Association rates were measured under continuous flow of 5 µl/minute using concentrations ranging from 5.0×10^{-8} to 8.0×10^{-7} M. k_{on} was determined from a plot of $(\ln(dR/dt))/t$ *versus* concentration (Karlsson *et al.*, 1991). To verify that differences in k_{on} were not due to differences in immunoreactivity, the relative concentrations of functional scFv were determined using SPR in a BIAcore (Karlsson *et al.*, 1993; Schier *et al.*, 1996b). Briefly, 4000 RU of c-erbB-2 ECD were coupled to a CM5 sensor chip and the rate of binding of C6.5 (RU/s) determined under a constant flow of 30 µl/minute. Over the concentration range of 1.0×10^{-9} M to 1.0×10^{-7} M, the rate of binding was proportional to the log of the scFv concentration. Purified scFv were diluted to the same concentration (1.0×10^{-8} M and 2.0×10^{-8} M) as determined by A_{280}. The rate of binding to c-erbB-2 ECD was measured and used to calculate the concentration based on the standard curve constructed from C6.5. Dissociation rates were measured using a constant flow of 25 µl/minute and a scFv concentration of 1.0×10^{-6} M. k_{off} was determined during the first two minutes of dissociation for scFv mutated in V_L CDR3 (Karlsson *et al.*, 1991) and during the first 15 to 60 minutes for clones with k_{off} below 5×10^{-4} s^{-1} (scFv mutated in V_H CDR3 and combined scFv). To exclude rebinding, k_{off} was determined in the presence and absence of 5.0×10^{-7} M c-erbB-2 ECD as described by Schier *et al.* (1996b).

Cell surface retention assay

The cell surface retention of selected scFv was determined on live SK-OV-3 cells using a fluorescence activated cell sorter (FACS). Purified scFv were labeled with NHS-LC-Biotin (Pierce) using the manufacturer's instructions. The concentration of immunoreactive-biotinylated scFv was calculated using SPR as described above and by Schier *et al.* (1996b). The efficiency of biotinylation was also determined in a BIAcore using a flow cell to which 5000 RU of streptavidin was coupled. The total responses after association were compared between samples and concentrations of scFv were adjusted using the results obtained from the BIAcore. For the assay, aliquots of SK-OV-3 cells (1.2×10^7 c-erbB-2 positive cells) were incubated with 14 µg biotinylated scFv in a total volume of 0.5 ml (1 µM scFv) FACS buffer (PBS containing 1% BSA and 0.1% NaN_3) for one hour at 37°C. Cells were washed twice with 10 ml FACS buffer (4°C) and resuspended in 12 ml FACS buffer and further incubated at 37°C. Aliquots of cells (0.5 ml from 12 ml containing 5×10^5 cells) were taken after five minutes, every 15 minutes for the first hour and after two hours repeating the wash and resuspension cycle. Washed cell aliquots were fixed with 1% paraformaldehyde, washed

twice with FACS buffer, and incubated for 15 minutes at 4°C with a 1:800 dilution of phycoerythrine-labeled streptavidin (Pierce). Fluorescence was measured by FACS and the percentage retained fluorescence on the cell surface plotted *versus* the time points. scFv used for the cell surface retention assay were C6.5 ($K_d = 1.6 \times 10^{-8}$ M), C6ML3-9 ($K_d = 1.0 \times 10^{-9}$ M), C6MH3-B1 ($K_d = 1.2 \times 10^{-10}$ M), and the anti-digoxin scFv 26-10 (Huston *et al.*, 1988) as negative control.

Construction of scFv combining higher affinity V_H and V_L genes

The V_L CDR3 gene sequences of the two highest affinity scFv isolated from the C6VLCDR3 library (C6ML3-9 or C6ML3-12) were combined with the highest affinity scFv previously obtained from light chain shuffling (C611, $K_d = 2.5 \times 10^{-9}$ M; Schier *et al.*, 1996b). The C6L1 plasmid (10 ng/μl) was used as a template for PCR amplification using primers LMB3 and either PML3-9 or PML3-12 (Table 1). The gel purified PCR fragments were reamplified using primers LMB3 and HuJλ 2-3ForNot (Marks *et al.*, 1991) to introduce a *Not*I restriction site at the 3'-end of the scFv. The gel purified PCR fragments were digested with *Nco*I and *Not*I and ligated into pUC119 Sfi-NotmycHis digested with *Nco*I and *Not*I. The resulting scFv were designated C6-9L1 and C6-12L1. The V_L genes of C6-9L1 and C6-12L1 were combined with the V_H genes of the two highest affinity scFv from the C6VHCDR3 libraries (C6MH3-B1 and C6MH3-B47). The rearranged V_H genes of C6MH3-B1 and -B47 were amplified by PCR using the primer LMB3 and PC6VH1FOR (Schier *et al.*, 1996b), digested with *Nco*I and *Xho*I (located in FR4 of the heavy chain) and ligated into C6-9L1 or C6-12L1 digested with *Nco*I and *Xho*I to create C6-B1L1 and C6-B47L1. The heavy chain of C6MH3-B1 or C6MH3-B47 was amplified by PCR using LMB3 and one of the PCD primers (PCD1, PCD2, PCD3, PCD5, or PCD6; Table 1) to construct combinations of scFv from the C6VHCDR3B and D libraries. The purified PCR fragments were spliced with the V_L fragment of C6ML3-9 (VHD2) that was used to create the C6VHCDR3D library exactly as described above. The full length scFv gene was digested with *Nco*I and *Not*I and ligated into pUC119 Sfi-NotmycHis. Clones were termed C6-B1D1, -B1D2, -B1D3, -B1D5, -B1D6, -B47D1, -B47D2, -B47D3, -B47D5, and -B47D6. Colonies were screened for the presence of the correct insert by PCR fingerprinting and confirmed by DNA sequencing. scFv were expressed, purified, and affinities determined by SPR, as described above.

High resolution functional scan of C6.5 V_H CDR3

A high resolution functional scan of the C6.5 V_H CDR3 was performed by individually mutating residues 95 to 99, 100a to 100d, and 100g to 102 to alanine. The pair of cysteine residues (100 and 100e) were simultaneously mutated to serine. Residue 100f (alanine) was not studied. Mutations were introduced by oligonucleotide directed mutagenesis using the method of Kunkel *et al.* (1987). Insertion of the correct mutation was verified by DNA sequencing, and scFv was expressed (De Bellis & Schwartz, 1990; Breitling *et al.*, 1991) and purified by IMAC (Hochuli *et al.*, 1988). Affinities were determined by SPR as described above (Schier *et al.*, 1996b) and compared to C6.5 scFv.

Modeling of location of mutations

The location of mutations in mutated scFv was modeled on the structure of the Fab KOL (Marquart *et al.*, 1980) using the program MidasPlus (Ferrin *et al.*, 1988) on a Silicon Graphics workstation.

Acknowledgements

We thank Gerald Apell for assistance in scFv purification, and James S. Huston for providing c-erbB-2 ECD and 26-10 scFv. This work was supported in part by the U. S. Army Medical Research, Development, Acquisition, and Logistics Command (Prov.) under grant no. DAMD17-94-J-4433, by the Arnold and Mabel Beckman Foundation, and by the CaPCURE Foundation.

References

Abergel, C. & Claverie, J.-M. (1991). A strong propensity toward loop formation characterizes the expressed reading frames of the D segments at the Ig H and T cell receptor loci. *Eur. J. Immunol.* **21**, 3021–3025.

Adams, G. P., DeNardo, S. J., Amin, A., Kroger, L. A., DeNardo, G. L., Hellstrom, I. & Hellstrom, K. E. (1992). Comparison of the pharmacokinetics in mice and the biological activity of murine 16 and human-mouse chimeric ch-16 antibody. *Antibody, Immunoconj. Radiopharm.* **5**, 81–95.

Adams, G. P., McCartney, J. E., Tai, M.-S., Oppermann, H., Huston, J. S., Stafford, W. F., Bookman, M. A., Fand, I., Houston, L. L. & Weiner, L. M. (1993). Highly specific *in vivo* tumor targeting by monovalent and divalent forms of 741F8 anti-c-erbB-2 single chain Fv. *Cancer Res.* **53**, 4026–4034.

Alzari, P. M., Spinelli, S., Mariuzza, R. A., Boulot, G., Poljak, R. J., Jarvis, J. M. & Milstein, C. (1990). Three-dimensional structure determination of an anti-2-phenyloxazolone antibody: the role of somatic mutation and heavy/light chain pairing in the maturation of an immune response. *EMBO J.* **9**, 3807–3814.

Barbas, C. F., Hu, D., Dunlop, N., Sawyer, L., Cababa, D., Hendry, R. M., Nara, P. L. & Burton, D. R. (1994). *In vitro* evolution of a neutralizing human antibody to human immunodeficiency virus type 1 to enhance affinity and broaden strain cross-reactivity. *Proc. Natl Acad. Sci. USA*, **91**, 3809–3813.

Betz, A. G., Neuberger, M. S. & Milstein, C. (1993). Discriminating intrinsic and antigen-selected mutational hotspots in immunoglobulin genes. *Immunol. Today*, **14**, 405–411.

Breitling, F., Dubel, S., Seehaus, T., Klewinghaus, I. & Little, M. (1991). A surface expression vector for antibody screening. *Gene*, **104**, 147–153.

Chothia, C. & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901–917.

Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn M. B., Winter, G. Structural repertoire of the human VH segments. *J. Mol. Biol.* **227**, 799–817, 1992.

Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature*, **352**, 624–628.

Clauss, M. A. & Jain, R. K. (1990). Interstitial transport of

rabbit and sheep antibodies in normal and neoplastic tissues. *Cancer Res.* **50**, 3487–3492.

Colcher, D., Minelli, F. M., Roselli, M., Muraro. R., Simpson-Milenic, D. & Schlom, J. (1988). Radioimmunolocalization of human carcinoma xenografts with B72.3 second generation monoclonal antibodies. *Cancer Res.* **48**, 4597–4603.

Cox, J. P. L., Tomlinson, I. M. & Winter, G. (1994). A directory of human germ-line V_κ segments reveals a strong bias in their usage. *Eur. J. Immunol.* **24**, 827–836.

Crothers, D. M. & Metzger, H. (1972). The influence of polyvalency on the binding properties of antibodies. *Immunochemistry*, **9**, 341–357.

Cunningham, B. C. & Wells, J. A. (1993). Comparison of a structural and a functional epitope. *J. Mol. Biol.* **234**, 554–563.

Davies, D. R., Padlan, E. A. & Sheriff, S. (1990). Antibody-antigen complexes. *Annu. Rev. Biochem.* **59**, 439–473.

De Bellis, D. & Schwartz, I. (1990). Regulated expression of foreign genes fused to lac: control by glucose levels in growth medium. *Nucl. Acids Res.* **18**, 1311.

Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucl. Acids Res.* **16**, 6127–6145.

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge R. (1988). The MIDAS display system. *J. Mol. Graphics*, **6**, 13–27.

Foote, J. & Eisen, H. N. (1995). Kinetic and affinity limits on antibodies produced during immune responses. *Proc. Natl Acad. Sci. USA*, **92**, 1254–1256.

Fujimori, K., Covell, D. G., Fletcher, J. E. & Weinstein, J. N. (1990). A modeling analysis of monoclonal antibody percolation through tumors: a binding site barrier. *J. Nucl. Med.* **31**, 1191–1198.

Gibson, T. J. (1984). Studies on the Epstein-Barr virus genome. PhD thesis, University of Cambridge.

Griffiths, A. D., Malmqvist, M., Marks, J. D., Bye, J. M., Embleton, M. J., McCafferty, J., Baier, M., Holliger, K. P., Gorick, B. D., Hughes-Jones, N. C., Hoogenboom, H. R. & Winter, G. (1993). Human anti-self antibodies with high specificity from phage display libraries. *EMBO J.* **12**, 725–734.

Gussow, D. & Clackson, T. (1989). Direct clone characterization from plaques and colonies by the polymerase chain reaction. *Nucl. Acids Res.* **17**, 4000.

Hawkins, R. E., Russell, S. J. & Winter, G. (1992). Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J. Mol. Biol.* **226**, 889–896.

Hawkins, R. E., Russell, S. J., Baier, M. & Winter, G. (1993). The contribution of contact and non-contact residues of antibody in the affinity of binding to antigen. The interaction of mutant D1.3 antibodies with lysozyme. *J. Mol. Biol.* **234**, 958–964.

Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R. & Stuber, D. (1988). Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Bio/Technology*, **6**, 1321–1325.

Hoogenboom, H. R. & Winter, G. (1992). Bypassing immunisation: human antibodies from synthetic repertoires of germ line V_H-gene segments rearranged *in vitro*. *J. Mol. Biol.* **227**, 381–388.

Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., Hudson, P. & Winter, G. (1991). Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucl. Acids Res.* **19**, 4133–4137.

Huston, J. S., Levinson, D., Mudgett, H. M., Tai, M. S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R. & Oppermann, H. (1988). Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **85**, 5879–5883.

Johnsson, B., Löfås, S. & Lindqvist, G. (1991). Immobilization of proteins to a carboxymethyldextran modified gold surface for BIAcore in surface plasmon resonance. *Anal. Biochem.* **198**, 268–277.

Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1991). *Sequences of Proteins of Immunological Interest*, US Department of Health and Human Services, US Government Printing Office.

Karlsson, R., Michaelsson, A. & Mattsson, L. (1991). Kinetic analysis of monoclonal antibody-antigen interactions with a new biosensor based analytical system. *J. Immunol. Methods*, **145**, 229–240.

Karlsson, R., Fagerstam, L., Nilshans, H. & Persson, B. (1993). Analysis of active antibody concentration. Separation of affinity and concentration parameters. *J. Immunol. Methods*, **166**, 75–84.

Kelley, R. F. & O'Connell, M. P. (1993). Thermodynamic analysis of an antibody functional epitope. *Biochemistry*, **32**, 6828–6835.

Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367–382.

Low, N., Holliger, P. & Winter, G. (1996). Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. *J. Mol. Biol.* In the press.

Lowman, H. B. & Wells, J. A. (1993). Affinity maturation of human growth hormone by monovalent phage display. *J. Mol. Biol.* **234**, 564–578.

Lowman, H. B., Bass, S. H., Simpson, N. & Wells, J. A. (1991). Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry*, **30**, 10832–10838.

Malmqvist, M. (1993). Surface plasmon resonance or detection and measurement of antibody-antigen affinity and kinetics. *Curr. Opin. Immunol.* **5**, 282–286.

Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D. & Winter, G. (1991). By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222**, 581–597.

Marks, J. D., Griffiths, A. D., Malmqvist, M., Clackson, T. P., Bye, J. M. & Winter, G. (1992). By-passing immunization: building high affinity human antibodies by chain shuffling. *Biotechnology (NY)*, **10**, 779–783.

Marks, J. D., Ouwehand, W. H., Bye, J. M., Finnern, R., Gorick, B. D., Voak, D., Thorpe, S., Hughes-Jones, N. C. & Winter, G. (1993). Human antibody fragments specific for blood group antigens from a phage display library. *Bio/Technology*, **11**, 1145–1149.

Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. (1980). Crystallographic refinement and atomic models of the intact immunoglobulin molecule KOL and its antigen binding fragment at 3.0 Å and 1.9 Å resolution. *J. Mol. Biol.* **141**, 369–391.

McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*, **348**, 552–554.

McCartney, J. E., Tai, M.-S., Hudziak, R. M., Adams, G. P., Weiner, L. M., Jin, D., Stafford, W. F., III, Liu, S., Bookman, M. A., Laminet, A., Fand, I., Houston, L. L., Oppermann, H. & Huston, J. S. (1995). Engineering disulfide-linked single-chain Fv dimers [(sFv')$_2$] with improved solution and targeting properties: anti-digoxin 26-10 (sFv')$_2$ and anti-c-erbB-2 741F8 (sFv')$_2$ made by protein folding and bonded through C-terminal cysteinyl peptides. *Protein Eng.* **8**, 301–314.

Mian, I. S., Bradwell, A. R. & Olson, A. J. (1991). Structure, function and properties of antibody binding sites. *J. Mol. Biol.* **217**, 133–151.

Milenic, D. E., Yokota, T., Filpula, D. R., Finkelman, M. A. J., Dodd, S. W., Wood, J. F., Whitlow, M., Snow, P. & Schlom, J. (1991). Construction, binding properties, metabolism, and targeting of a single-chain Fv derived from the pancarcinoma monoclonal antibody CC49. *Cancer Res.* **51**, 6363–6371.

Miller, J. H. (1972). *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Nieba, L., Krebber, A. & Plückthun, A. (1996). Competition BIAcore for measuring true affinities: large differences from values determined from binding kinetics. *Anal. Biochem.* **234**, 155–165.

Novotny, J., Bruccoleri, R. E. & Saul, F. A. (1989). On the attribution of binding energy in antigen-antibody complexes McPC 603, D1.3, and HyHEL-5. *Biochemistry*, **28**, 4735–4749.

Reynaud, C.-A., Garcia, C., Hein, W. R. & Weill, J.-C. (1995). Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell*, **80**, 115–125.

Riethmueller, G., Schneider-Gädicke, E., Schlimok, G., Schmiegel, W., Raab, R., Höffken, K., Gruber, R., Pichlmaier, H., Hirche, H., Pichlmayr, R., Buggisch, P., Witte, J. (1994). Randomised trial of monoclonal antibody for adjuvant therapy of Dukes' C colorectal carcinoma. German Cancer Aid 17-1A study group. *Lancet*, **343**, 1177–1183.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1990). *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Schier, R. & Marks, J. D. (1996). Efficient *in vitro* affinity maturation of phage antibodies using BIAcore guided selections. *Human Antibodies and Hybridomas*. In the press.

Schier, R., Marks, J. D., Wolf, E. J., Apell, G., Wong, C., McCartney, J. E., Bookman, M. A., Huston, J. S., Houston, L. L., Weiner, L. M. & Adams, G. P. (1995). *In vitro* and *in vivo* characterization of a human anti-c-erbB-2 single-chain Fv isolated from a filamentous phage antibody library. *Immunotechnology*, **1**, 73–81.

Schier, R., Balint, R. F., McCall, A., Apell, G., Larrick, J. W. & Marks, J. D. (1996a). Identification of functional and structural amino acid residues by parsimonious mutagenesis. *Gene*, **169**, 147–155.

Schier, R., Bye, J., Apell, G., McCall, A., Adams, G. P., Malmqvist, M., Weiner, L. M. & Marks, J. D. (1996b). Isolation of high affinity monmeric human anti-c-erbB-2 single-chain Fv using affinity driven selection. *J. Mol. Biol.* **255**, 28–43.

Sharkey, R. M., Gold, D. V., Aninipot, R., Vagg, R., Ballance, C., Newman, E. S., Ostella, F., Hansen, H. J. & Goldenberg, D. M. (1990). Comparison of tumor targeting in nude mice by murine monoclonal antibodies directed against different human colorectal antigens. *Cancer Res.* **50**, 828s–834s.

Sharon, J. (1990). Structural correlates of high antibody affinity: three engineered amino acid substitutions can increase the affinity of an anti-p-azophenylarsonate antibody 200-fold. *Proc. Natl Acad. Sci. USA*, **87**, 4814–4817.

Tomlinson, I. M., Cox, J. P. L., Gherardi, E., Lesk, A. M. & Chothia, C. (1995). The structural repertoire of the human V_κ domain. *EMBO J.* **14**, 4628–4638.

Tomlinson, I. M., Walter, G., Jones, P. T., Dear, P. H., Sonnhammer, E. L. L. & Winter, G. (1996). The imprint of somatic hypermutation on the repertoire of human germline V genes. *J. Mol. Biol.* **256**, 813–817.

Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., McCafferty, J., Hodits, R. A., Wilton, J. & Johnson, K. S. (1996). Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nature Biotechnol.* **14**, 309–314.

Wagner, S. D., Milstein, C. & Neuberger, M. S. (1995). Codon bias targets mutation. *Nature*, **376**, 732.

Williams, S. C. & Winter, G. (1993). Cloning and sequencing of human immunoglobulin V_λ gene segments. *Eur. J. Immunol.* **23**, 1456–1461.

Yamada, M., Wasserman, R., Reichard, B. A., Shane, S., Caton, A. J. & Rovera, G. (1991). Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult and peripheral blood B lymphocytes. *J. Exp. Med.* **173**, 395–407.

Yang, W. P., Green, K., Pinz-Sweeney, S., Briones, A. T., Burton, D. R. & Barbas, C. F., III (1995). CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. *J. Mol. Biol.* **254**, 392–403.

Yokota, T., Milenic, D., Whitlow, M. & Schlom, J. (1992). Rapid tumor penetration of a single-chain Fv and comparison with other immunoglobulin forms. *Cancer Res.* **52**, 3402–3408.

Edited by J. Karn

(Received 15 May 1996; received in revised form 12 August 1996; accepted 23 August 1996)

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.